Twitter Sentimental Analysis
ML Course Project

P Charishma Kumari(M20AIE270)
Venkata Lakshmi Sunayana(M20AIE311)

Execution Procedure:

Github link:
https://github.com/charishmacherry/ML_CourseProject

Steps:
1. Download the .ipynb file
2. Download the Kaggle.json file
3. Upload the .ipynb file and .json file into collab.
4. Execute the cells.

Sentimental Analysis Model:

Here we are considering traditional machine learning models as follows for sentimental analysis:

- Guasian Naive Bayes Classification
- Logistic Regression

We have used scratch technoque to preprocess the text.
After preprocessing and applying the progress using progress_apply the progress bar that we get is as follow:

```
/usr/local/lib/python3.7/dist-packages/tqdm/std.py:658: FutureWarning: The Panel class is removed from pandas. Accessing it from the top-level namespace will also be rem
  from pandas import Panel
100% 1600000/1600000 [00:46<00:00, 34330.94it/s]
```

	sentiment	text	pre_text
0	0	@switchfoot http://twitpic.com/2y1zl - Awww, t...	USER URL aww that s a bummer you shoulda ...
1	0	is upset that he can't update his Facebook by ...	is upset that he can t update his facebook by ...
2	0	@Kenichan I dived many times for the ball. Man...	USER i dived many times for the ball managed...
3	0	my whole body feels itchy and like its on fire	my whole body feels itchy and like its on fire
4	0	@nationwideclass no, it's not behaving at all....	USER no it s not behaving at all i m mad w...

a. Guassian Naive Bayes Classification:

```
Accuracy: 0.804725
Confusion Matrix:
 [[32453  7547]
 [ 8075 31925]]
Classification Report:
              precision    recall  f1-score   support

    NEGATIVE       0.80      0.81      0.81     40000
    POSITIVE       0.81      0.80      0.80     40000

    accuracy                           0.80     80000
   macro avg       0.80      0.80      0.80     80000
weighted avg       0.80      0.80      0.80     80000
```

b. Logistic Regression:

```
Accuracy: 0.81895
Confusion Matrix:
 [[32407  7593]
 [ 6891 33109]]
Classification Report:
              precision    recall  f1-score   support

    NEGATIVE       0.82      0.81      0.82     40000
    POSITIVE       0.81      0.83      0.82     40000

    accuracy                           0.82     80000
   macro avg       0.82      0.82      0.82     80000
weighted avg       0.82      0.82      0.82     80000
```

References:

[1] Dataset : https://www.kaggle.com/datasets

[2] Guasian Naive Bayes: https://towardsdatascience.com/learning-by-implementing-gaussian-naive-bayes-3f0e3d2c01b2
https://www.analyticsvidhya.com/blog/2021/01/gaussian-naive-bayes-with-hyperpameter-tuning/

[3] Logistic Regression: https://www.analyticsvidhya.com/blog/2021/03/logistic-regression/