**README**

**Group Information:** This is the final project for BUDT 737 and our topic is Pandemic Tweet Analysis. Our group members include: Yudong Lu, Fengyi Zhao, Javan Reuto, and Bradley R. [Mascarenhas](https://umd.instructure.com/courses/1314706/users/4280535).

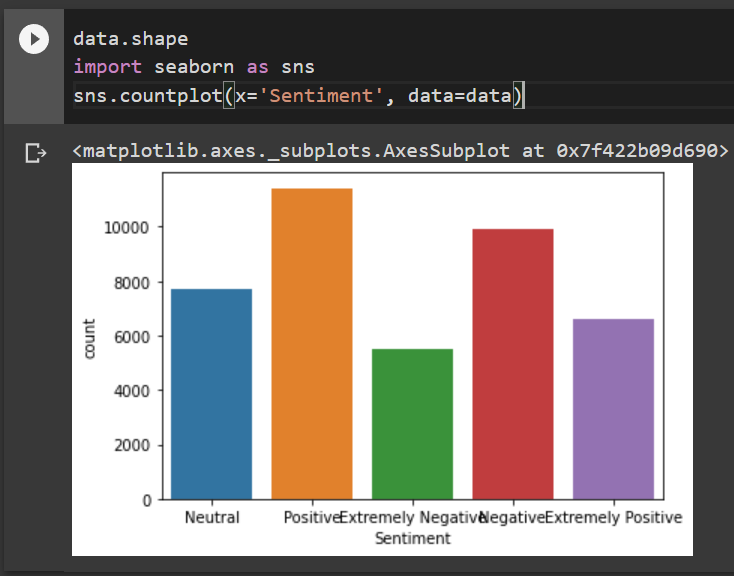
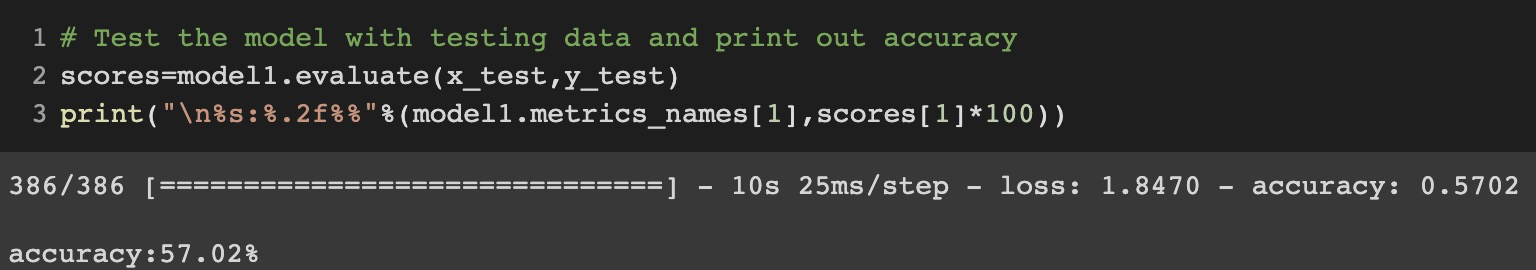
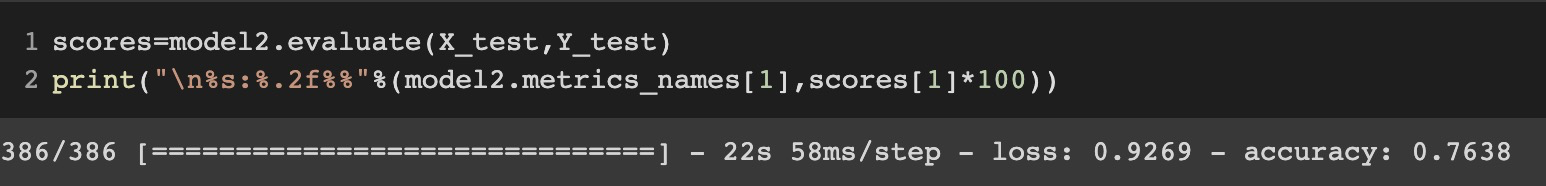
**Background Information:** The project is to identify the nature of the Covid Tweets pulled from Twitter during the coronavirus period based on Natural Language Processing.

**File Contains:** We have the following four files for our project:

* **Dataset:** Could be downloaded from [Pandemic Tweet Challenge | Kaggle](https://www.kaggle.com/c/pandemic-tweet-challenge/overview), we only used the train dataset for our project
* **737\_final\_project.ipynb:** This file contains all the codes of our project and includes detailed report
* **Project 737 Report.pdf:** This file is a pure report part of the project
* **README.doc:** An explanatory file of how to run the code and output summary

**How to run the code:** The best way to run the code is to open the jupyter notebook via colab. First download the dataset from the above link and upload it to google doc. Then change the path to read the dataset. You could also download the dataset to the local and run it through your terminal and navigate to the directory where you would like to save your notebook. We recommend running it through colab.

**Output:**

* **Visualization:** We did a simple bar chart for inspection.****
* **Fully Connected Feedforward Network Output:** We receive accuracy of 57.02%. ****
* **LSTM Output:** We receive accuracy of 76.4%****

**Potential problems:** Some tweet sentiments contained hashtags and our project did not take the hashtags into consideration. We simply remove the hashtag sign(#) but did not remove the hashtag words. Therefore some tweets might contain meaningless words. Besides that we did not find other problems.