**LSML2 Final Project Read Me**

1. **Project documentation**

In this project I built the Twitter sentiment analysis service.

I used pre-trained BERT model and fine-tuned it using Kaggle - Sentiment140 dataset with 1.6 million tweets.

This service is deployed as docker container with Flask web server inside which serves a web page where you can paste a tweet and get the prediction on tweet’s sentiment.

To improve performance, this service can be deployed on multiple servers behind a load balancer. This is a synchronous request-response service, it does not use queue and it does not have a database.

In order to run this service, you need to run the docker container.

docker run -p 5000:5000 sm:latest (if you built it yourself following the instructions below)

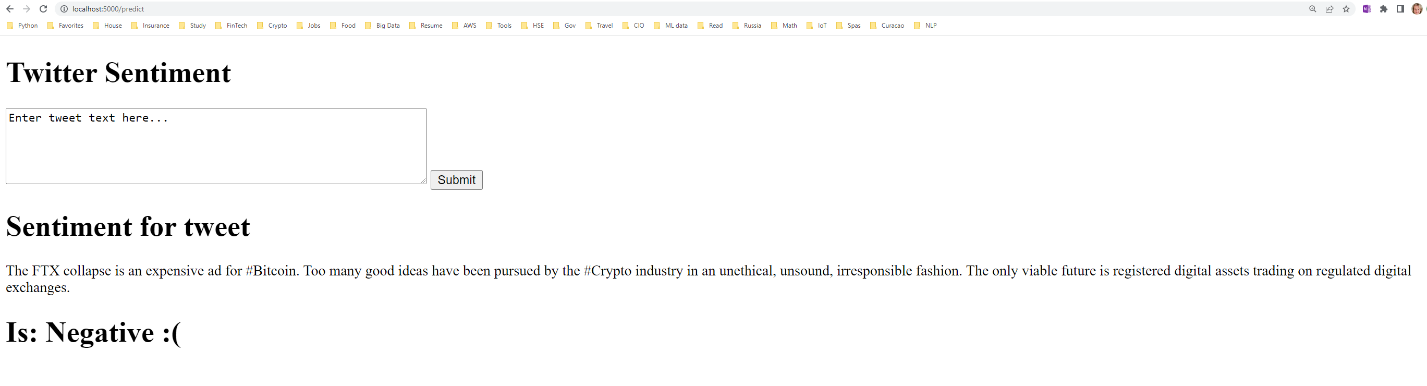
or

docker run -p 5000:5000 susanmalkin/lsml2\_sm:latest (if you want to pull it from DockerHub repository)

This container will launch Flask web server on local host port 5000. Use this link in your browser <http://localhost:5000/> to access this service.



Cut and paste a message from Twitter into the box and hit Submit button.



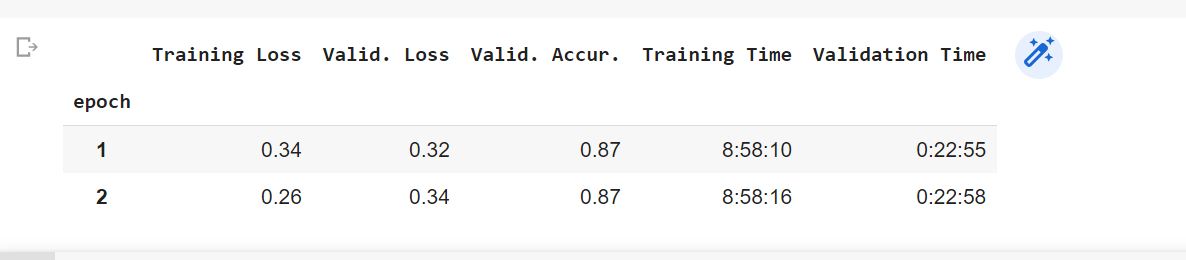
1. **Model and Dataset**

I used pre-trained BERT model:[bert-base-uncased · Hugging Face](https://huggingface.co/bert-base-uncased?text=Paris+is+the+%5BMASK%5D+of+France) with Kaggle dataset: <https://www.kaggle.com/datasets/kazanova/sentiment140>

Model training code is in LSML2\_final\_project\_Susan\_Malkin.ipynb file. This program needs access to training.tweet.csv file from the Kaggle site to load dataset. I don’t use the entire data set. I used 200 records for test set and 600,000 for train/validation set (90-10 split).

During execution this program will create directories called 1111model<epoch\_number> to store model state after each epoch.

These are result of training and testing:



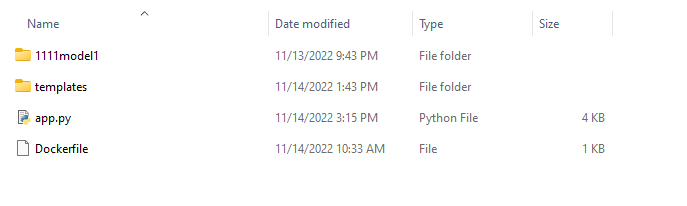




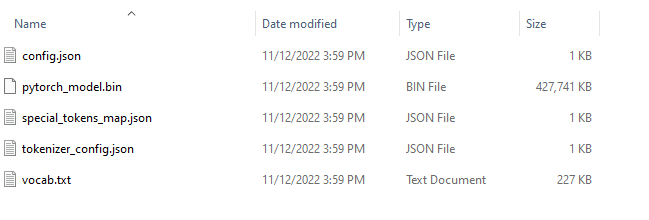
It looks like the model tends to overfit with large number of epochs. I used 2 epochs for training. You need directory 1111model1 and its content to create Docker container.

1. **Service Deployment**

Create a project directory with the following structure:



Where 1111model1 is directory created during model training with its content.



And templates directory contains home.html file.

Run

docker build -t sm .

command in this directory followed by

docker run -p 5000:5000 sm:latest