Pongsasit Thongpramoon

Machine learning Engineer Nanodegree Capstone Project Proposal.

I will describe this proposal by describing the 7 key points.

1. The project's **domain background** — the field of research where the project is derived; Sentimental Analysis for classify people with depression out from normal person. Using deep learning.
2. A **problem statement** — a problem being investigated for which a solution will be defined;
    1. Problem
        a. Depression is a mental illness that is not taken seriously in some countries.
        b. Depression is the leading cause of disability worldwide. Almost 75% of people with mental disorders remain untreated in developing countries with almost 1 million people taking their lives each year. In addition, according to the World Health Organization (WHO), 1 in 13 globally suffers from anxiety. The WHO reports that anxiety disorders are the most common mental disorders worldwide with specific phobia, major depressive disorder and social phobia being the most common anxiety disorders.
        c. Earlier we can detect the disease, the easier to make a treatment.
    2. Solution
        a. Make depression detector using sentimental analysis.
3. The **datasets and inputs** — data or inputs being used for the problem;

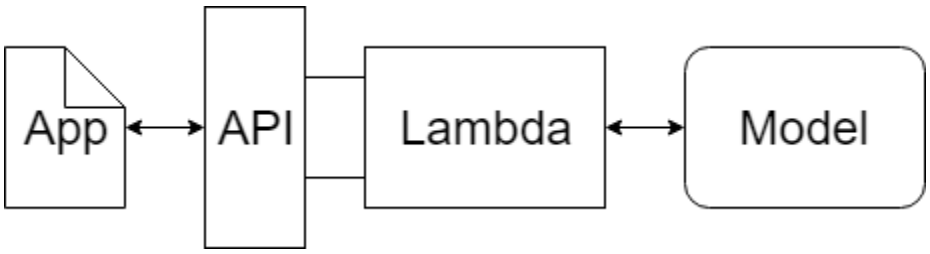
    Dataset 1: Sentiment140

    https://www.kaggle.com/kazanova/sentiment140

    Dataset 2: Web scraping using TWINT for depressive tweets.

4. A **solution statement** — the solution proposed for the problem given;

    1. Make this project to production



    I will deploy model to production scale using AWS Lambda, API Gateway, Sagemaker endpoint. So, everyone can access to this application even though they don't know any basic of coding.

    2. Text Processing
        a. TWINT for depressive data
        b. Decapitalize the alphabets
        c. Remove punctuation mark
        d. Remove common word
    3. Feature Extraction
        a. Bag of words
        b. TF-IDF
        c. One-hot encoding
        d. Word embedding
        e. Word2Vec
        f. Glove
    4. Modeling
        a. Simple RNN model

   b. RNN with LSTM model
   c. AWS BlazingText algorithm
   d. ULMFiT

5. A **benchmark model** — some simple or historical model or result to compare the defined solution to;

   AWS BlazingText algorithm

6. A set of **evaluation metrics** — functional representations for how the solution can be measured;
   a. Accuracy
   b. Precision
   c. Recall
   d. F-score
7. An outline of the **project design** — how the solution will be developed and results obtained.
   - As you can see in Fig.1

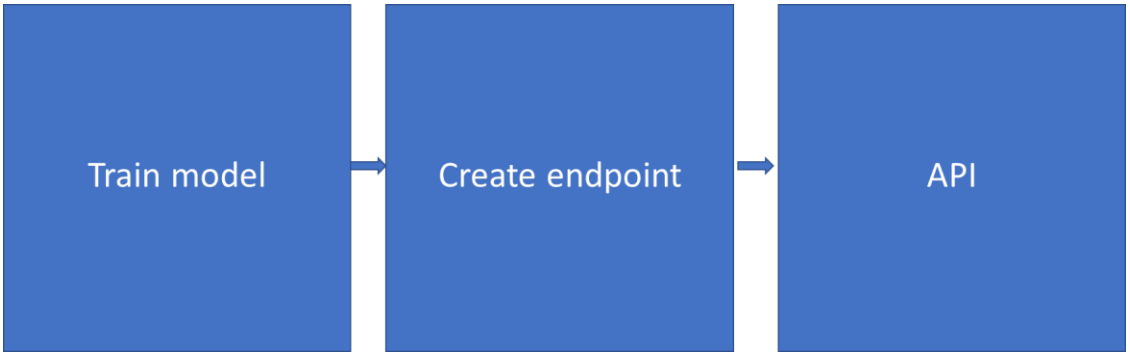


Fig1. Project design

Will train and evaluate model on AWS Sagemaker instance notebook. Then use only the best model for deployment

- Code will be written by PyTorch.

Will create an endpoint then use AWS Lambda and API Gateway to make API