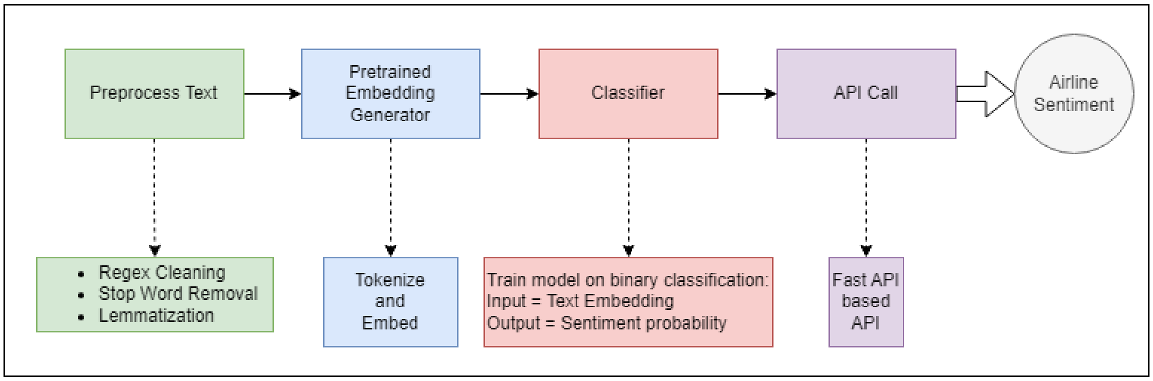
Name: Tashvik Dhamija
Email: dhamijatashvik@gmail.com
GitHub Repo for project: https://github.com/TashvikDhamija/TrueFoundry-ML-Internship-Project

Sentiment Analysis API

1. Pipeline

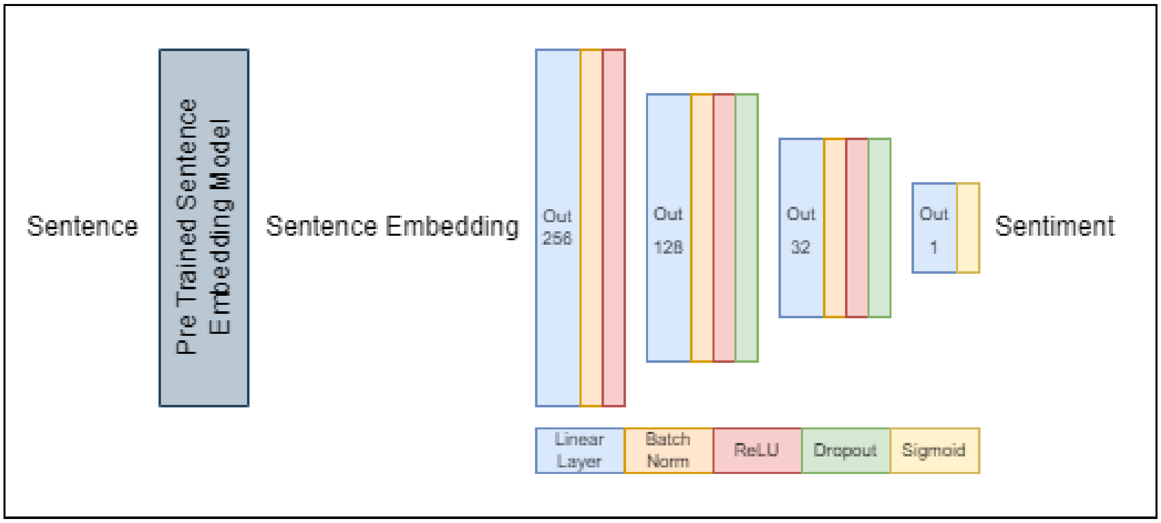


The pipeline to solve the given problem has been described in the above figure. The data points are firstly processed as mentioned. A pre-trained sentence embedding generator model is loaded. It is better than creating a model to generate embeddings due to the small size of the dataset. An MLP classifier is trained to perform binary classification on these sentence embeddings. Then, an API is created to query the embedding model and the classifier for public use of the created pipeline. Experiments and Training details are given in the following sections.

2. Model Details

The given dataset has been split into a 3:1 ratio for training and test purposes. The batch size, dropout rate, learning rate, and epochs are taken as 8, 0.2, 0.01, and 100 respectively. An Adam with Adagrad optimizer is used to converge a Binary Cross Entropy Loss

The following classifier is used:



3. Experiments

Pretrained Model	Test Accuracy	Speed (from docs)	Size (from docs)
paraphrase-MiniLM-L3-v2	91.7533%	19000	61
all-MiniLM-L6-v2	91.4414%	14200	80
all-distilroberta-v1	92.2037%	4000	290
all-mpnet-base-v2	91.9958%	2800	420

The above experiments show that all models give a relatively similar performance on the data. Hence ‘paraphrase-MiniLM-L3-v2’ is the model since it is the smallest and the fastest model without losing a ton of performance.

4. Video Link https://drive.google.com/file/d/1DPFkMssEZk02897wCgA3XwEwwEn9AouG/view?usp=sharing