Initial Screen

Add Comment Here

Acceptable Comment i.e. Non-Toxic

VERY NICE, LOVED IT .....

Posted 10:10am 2 likes

Toxic Comment

YOURE SUCH A ****** , I HATE YOU ****

Removed 10:10am
This does not follow our community guidelines.
Please try again

WHY ?

Nowadays, more and more social media applications are leaning towards providing a safer and more all-inclusive environment for its users. A lot of backlashes was faced by the social media applications for explicit images, videos, and toxic comments towards the user. Hence, according to an article from social media today, Instagram is rolling out new moderation tools which will use machine learning to detect offensive language and spam. This could also mean that comments that are considered discouraging, ill-intentional, or outright hateful towards any user is being detected and removed. Using Machine Learning, I have tried to implement a Toxic Comment Classifier using a data set that contains examples of both toxic, and non-toxic comments. Using tokens that can preempt a comment being toxic or not, I am trying to increase the accuracy of an already existing project taken from https://www.kaggle.com/watermasterz/toxic-comments-classification/notebook.

HOW ?

This project makes use of four different models and compares the four different models' accuracy on how the comments have been classified. It uses neural networks, and I will experiment with different values of parameters and by changing the configurations of neural networks using the knowledge from the course, that will make the model more accurate.

Based on these experiments that I will try; I will construct a model that is a more accurate model using the same methods that this project has used.

The implementation contains 4 main types of neural network layers and uses different configuratiosn to make up the 4 models up for configuration. They are :

- LSTM Layers
- Densely Connected NN Layers / Fully Connected NN Layers
- Embedding Layers
- 1-D Convolutional Layers

CHALLENGE FACED AND IMPLEMENTATION

With an example that already has an accuracy of 95% is a challenge in itself as it is already a good accuracy score. Anything above this accuracy score can only be achieved with a lot of experiments, trials and efforts. So the major challenge for me is to increase the accuracy from this starting point. Other challenges also included understanding the primary concepts behind the already existing code and the insights behind it as well.

There were so many combinations that I had to try to get the most optimal combination to get the best possible accuracy for the combination I chose. Starting with an example that already has such high accuracy, I have to rely on using experiments and trials to obtain more accuracy. Using the insights from the course and articles referenced at the end of the blog, I was able to conduct about 23 experiments and use them to obtain a better accurate model. I was able to take this model from 95% to 99.5% approximately.