

Over view of files in this repository

1. **The Bert folder** contains the base model on which we first train our data.

The process is saved in a jupyter-notebook and can be loaded and run freely after installing the prerequisite libraries (such as transformers, pytorch and so on...)

In this file you can see how we loaded the data, presented it, analyzed it. After a few adjustments to the dataset, we load the distilBert model, the Tftrainer and perform fine-tuning. At the end, we can see the evaluation of the model, the confusion matrix results that we ran on the test_dataset.

2. Getting ready to create our own SpamGAN model

- **GPT2 text generator**

This folder contains the code for utilizing the GPT2 model for artificial text generation

- **GPT2 text-augmentations (fine-tuning)**

After a few tests we realized the GPT2 model was creating text that seems more like a “story” than actual SMS messages. We then decided to fine-tune it on our database.

Create a model and load it into the application.

based on this github repository:

<https://github.com/prakhar21/TextAugmentation-GPT2>

3. **SpamGAN folder**

Contains 2 files:

1. Classification_Algorithm_Comparisons – compare different supervised algorithms for our task and outputs a folder of pretrained models.
2. Dataset_Generation_With_Label – uses the fine-tuned GPT, the new models. Generates new text while classifying it with the different algorithms. If the classification matches the real label, the new sentence is then added to the dataset.