***Classification project***

**Problem**: A Social Media Monitoring company is tasked to sift through tweets and classify them as relevant (genuine tweet) or irrelevant (troll tweet) to help their Client focus on important tweets and ignore irrelevant ones.

**Solution**: Use Twitter data to build a Classification model using NLP

**Datasets**: https://www.kaggle.com/dataturks/dataset-for-detection-of-cybertrolls

***Classification project***

**Problem**: A Social Media Monitoring company is tasked to sift through tweets and classify them as positive, neutral or negative in sentiment in order for their Client to focus on what they are doing right, and what they are doing wrong.

**Solution**: Use Twitter and Facebook data to build a Sentiment Analysis model using NLP

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

***Translation project***

**Problem**: Translation Services Inc. has hired me to build an NLP model to automatically translate documents for Government entities.

**Solution**: I will use a parallel corpus French-English that is based on the European Parliament’s sessions. This data set can be found on the LionBridge website.

**Datasets**: <http://www.statmt.org/europarl/>

<https://github.com/ChalamPVS/Machine-Translation-Keras>

<https://stackabuse.com/python-for-nlp-neural-machine-translation-with-seq2seq-in-keras/>

<https://nlp.stanford.edu/projects/glove/> glove.6B.zip

<https://github.com/lukysummer/Machine-Translation-Seq2Seq-Keras/blob/master/machine_translation_seq2seq.py>

1. Did this Google search: data set for nlp translation english to french
2. The search provided this first result:

<https://lionbridge.ai/datasets/25-best-parallel-text-datasets-for-machine-translation-training/>

1. This brought me to: <http://www.statmt.org/europarl/> where I downloaded **parallel corpus French-English, 194 MB, 04/1996-11/2011**
2. This provided me with a French to English data set of European Parliament discussions: fr-en.tgz