30/09/2019

To do list:

1. Improve on relevance classifier
   1. Fine-tune models and compare
2. Categorizer classifier (maybe no longer needed)
3. Holdout set establishment
4. Checking on training set
   1. Balancing issues
5. Integrate these <http://www.crisislex.org/crisis-lexicon.html>
   1. Categories
   2. Disasters lexicon

Bonus round

1. Containerize full infrastructure
   1. Data extraction and ingestion pipeline
      1. Kafka - S3
   2. ML Infra
      1. Online Learning capable
   3. Front-end layer
      1. ?????
2. Github
3. Medium post

**Fine-tuning tutorial**

<https://mccormickml.com/2019/07/22/BERT-fine-tuning/>