Required Activity 6.4

Here are the results of my search on Kaggle and other sites for machine learning competitions which I believe are related to the financial regulator space, either directly or indirectly.

**Bengali.AI Speech Recognition**

Link: <https://www.kaggle.com/competitions/bengaliai-speech>

The aim of this competition is to improve the accuracy of speech recognition of Bengali. The benefit to the host organisation is hopefully a massively improved speech recognition model for one of the worlds most spoken languages. The current available models have word error rates as high as 74% (Google Speech API). This would be very beneficial to the financial regulatory market due to the massive economic growth in South Asia.

The unique application of machine learning here lies in the dataset which is offered for training. It is the first massively crowdsourced Bengali speech dataset and includes the variations in pace and tone with which current models struggle. This should allow for the training of a more diverse and accurate model.

My interest lies in the transcription capabilities a model like this would allow. In my industry, there are constant requests for communications in new languages to be monitored. The quality of the transcription impacts the quality of the alerts which are then output when surveilling these communications. I am considering entering to better my understanding of speech recognition.

**Natural Language Processing with Disaster Tweets**

Link: <https://www.kaggle.com/competitions/nlp-getting-started>

I understand the purpose of this competition is as an introduction to Kaggle competitions, so I won’t speak much to the benefit of the host. The main benefit here is to get more people involved in these competitions, whet their appetites, and increase the platform usage. The uniqueness of this application of ML is in the specific detection of false emergencies in tweets.

I’m very interested not only in the false emergencies as I believe this is a transferable knowledge/skill set directly to the financial regulatory products I work on. These products create alarms from the communications which are passed through them, applying rule-based surveillance logic. The model also pulls data from twitter to analyse, another useful tool in detecting fraudulent or untoward behaviour in communications. If all of twitter is screaming that the market is crashing, but traders are discussing increasing their long position, this may be considered suspicious and require investigation.

**The 3rd YouTube-8M Video Understanding Challenge**

Link: <https://www.kaggle.com/competitions/youtube8m-2019>

The goal of this competition was to create a model which would allow for temporal localization of topics with video. The host is Google Research and as the title suggests, this would have massive benefits for their sister branch, YouTube. Instead of allowing users to search just for video titles, tags, or descriptions, they could search for moments within a video, as recognized by the machine learning model. Recognition of localized video sections for the purposes of search and discovery is a great and unique use of machine learning.

Unfortunately, this competition is now closed as I would have loved to have participated. Not only from a personal interest in the project, but also because the opportunity to work with Google has long been a career goal of mine. I would definitely consider entering competitions like this in the future after gaining as much as I can from this course, of course.

**Skills Required**

I feel that the very competitive nature of any of these competitions, due to the size of the prizes offered, would require a keen aptitude in machine learning theory and implementation across the board. Having not looked at the datasets, it would be bold to assume they are clean enough to require no work, so data cleaning and manipulation would be key to start with.

A strong knowledge of python and the necessary libraries would no doubt be essential to be able to build and train the best model out in the competition. Being able to test and evaluate the models using any number of the metrics discussed so far in this course would be needed to pick and refine the models. A broad knowledge of the types of models and the different machine learning techniques would allow for a more balanced and considered approach which would save wasted development time.

Other, not as exciting, skills required could include: time management, teamwork/experience with collaboration tools, and persistence.