Title: Pandemic prediction

Author: Julie Pone

Goal: Use Machine Learning to predict Pandemic Waves

1. Use COVID infection data in US 2019~ 2020 as training data set.
2. Find out the 1st wave peak time.
3. Generate prediction model.
4. Use 2021 data set as test data set to verify prediction model for 2nd wave.
5. Retrain model using 2021 data set and predict the 3rd wave time in the future.

Variables:

* Gathering effect (Holiday, Lockdown….)
* Intervene Factor (Mask wearing, contact tracing…)
* Curd Community
* Vaccination rate

Data Source: Kaggle

**Capstone Proposal Guideline**

1. What is the problem you want to solve? Why is it an interesting problem?
2. What data are you going to use to solve this problem? How will you acquire this data?
3. In brief, outline your approach to solving this problem. You might not know everything in advance, and this approach may change later. This might include information like:
   1. a.Is this a supervised or unsupervised problem?
   2. b.If supervised, is it a classification or regression problem?
   3. c.What are you trying to predict?
   4. d.What will you use as predictors?
   5. e.Will you try a more “traditional” machine learning approach, a deep learning approach, or both?
4. What will be your final deliverable? Will it be an application deployed as a web service with an API or a more robust web/mobile app.
5. What computational resources would you need at a minimum to do this project? ​You may not have a very clear sense now but work with your mentor to come to an estimate. In real industrial applications, you’ll often be called upon to provide resource estimates at the beginning of a project.
   1. Processing power (CPU)
   2. Memoryc.Specialized hardware such as GPUs

Project Title: US Pandemic Wave Prediction for COVID-19

Goal: Use Machine Learning to predict COVID-19 Pandemic Waves in US

Data Source: Kaggle

Approach:

1. Use COVID infection data in CA 2019~ 2020 as training data set.
2. Find out the 1st wave peak time.
3. Generate prediction models using Time Series Prediction, LSTM, Deep Learning, Baysian, SVM…...
4. Compare MSE to see which model best fit.
5. Use 2021 data set as test data set to verify prediction model for 2nd wave.
6. Retrain model using 2021 data set and predict the 3rd wave time in the future.
7. Expand the prediction to US data

Variables:

* Gathering effect (Holiday, Lockdown….)
* Intervene Factor (Mask wearing, contact tracing…)
* Curd Community
* Vaccination rate

This project will use supervised regression model, starting from traditional and then deep learning approach.

The final product will be web demo to show peak predication and real peak occurrence comparison.

Statistic and computation resource will be needed. GPU is preferred.