**Project 2. Predicting Covid-19 Confirmed Cases Per Million People Based on Month Using Hidden Markov Model (HMM)**

You can finish this project in a group of 2.

1. Dataset

Attached “covid\_data.csv” is the dataset you need. The dataset contains the following data:

1. Date MM/DD/YYYY
2. New cases per million: daily new cases per million people. You will need to add this up to the entire month worth of new cases per million.
3. Total cases per million: total case so far instead of new cases. Use this optional feature if needed.
4. New cases 7 day average: Use this optional feature if needed.
5. Build your markov model. This one can be done by hand, but need to include your model in the project report.

You should have 10 hidden variables representing the level of infection: New cases per million < 1k; 1k – 2K; 2K – 3K; 3K – 4K; 4K – 5K; 5K – 6K; 6K – 7K; 7K – 8K; 8K – 9K; > 9 K.

The observable evidence should be month of the year.

1. Implement your HMM in any language you want. You have to code your HMM calculations from the ground up and not use existing HMM libraries. Your code should be able to do the following:
2. Given a month number 1- 12, filter the level of infection for the current month.
3. Given a month number 1- 12, predict the next 3 month level of infection.
4. Given a month number 1 - 12, Use Verbati algorithm to Explain the most likely level of infection from January until the input month.

Report Format:

1. Visualize your HMM
2. Screen shot of example input and output
3. Discuss the following:
   1. What’s the accuracy of your model
   2. Explain possible reasons why your model is not accurate
   3. Possible ways to improve your model accuracy.

What to turn in:

1. All code
2. Report