Contributors: Adithya Chandrasekar, Amrutha Vaidyam, Pranesh Monda, Aditi Anand, Christin Huene, Ongshu Dutta, Shreya Krishnan, Peyton Day, Josh Mansky, Tyler Sivertsen Overview

Predictive analytics are an important tool in managing pandemics. As our recent experience with the SARS-CoV-2 pandemic showed, the ability to predict trends in infection rates at a local level can greatly help in planning the response. In this project, we propose a method for predicting the 7-day average of new COVID-19 cases and the positivity rate based on historical public health data.

Methods

Since the data is in the form of a time series, we chose Long-Short Term Memory (LSTM) neural networks as the basis of our method. LSTMs are recurrent neural networks that are particularly well-suited to processing data that is sequential in nature. LSTMs are designed to handle both long-term dependencies (between inputs that occur far apart in time) and short-term dependencies. LSTMs consist of a repeating sequence of building blocks, each of which is composed of an input gate, a forget gate and an output gate.

We implemented our model using the PyTorch and scikit-learn machine learning frameworks. We first normalized the data using the MinMaxScaler module in scikit-learn. We used a 2-layer LSTM network with a hidden state dimension of 32 and trained it for 100 epochs with a learning rate of 0.01. Two instances of the model were used to predict the daily new cases and positivity rate, respectively. The 7-day average was computed by post-processing the daily predictions.

Results

I. <u>Daily New Cases</u>
   A. Training
      Figure 1 reports the normalized training loss of the model for new cases, which converges to a Mean Squared Error (MSE) of 0.038.

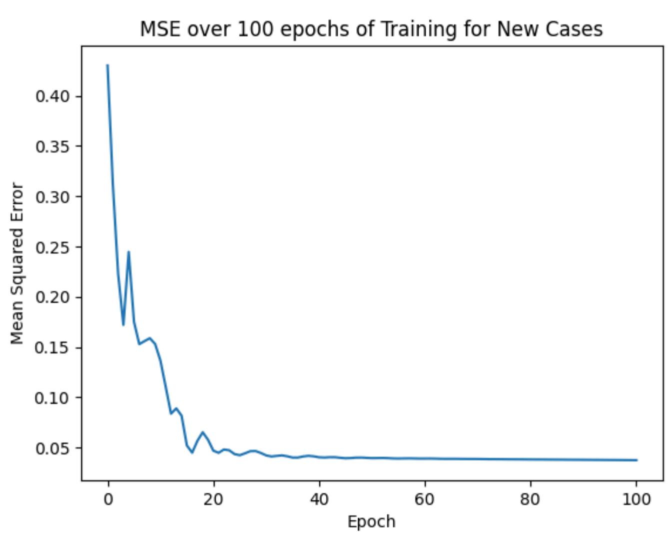

Figure 1. *MSE loss of the model predicting the daily new cases over the 100 epochs of training*

B. Scoring

Table 1 reports the prediction accuracy metrics for new cases for the period Aug 30, 2021 - Sept 26, 2021.

Table 1. *Root Mean Squared Error & Mean Absolute Error of the proposed model for new cases between Aug. 30, 2021 and Sept. 26, 2021*

Root Mean Squared Error	Mean Absolute Error
28931.32	15813.73

Figure 3. Model Prediction and Actual New Cases

II. Positivity Ratio

A. Training

Figure 3 reports the normalized training loss of the model for positivity ratio, which converges to a Mean Squared Error (MSE) of 6.42×10^{-5}.

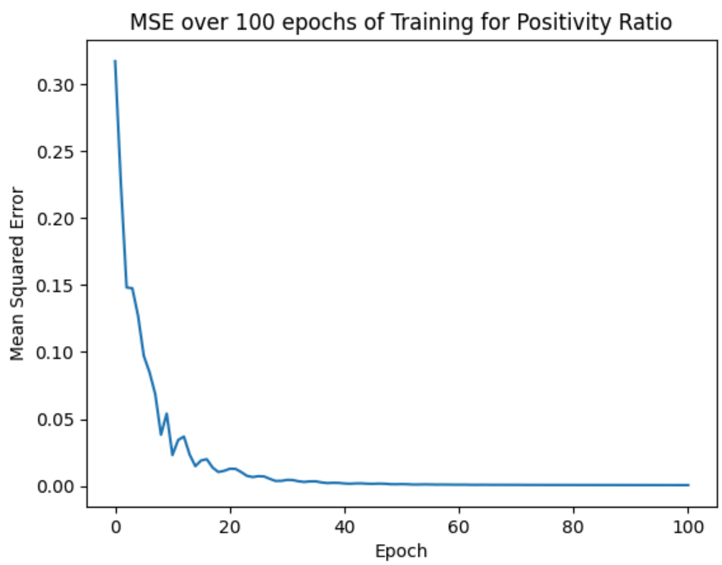

Figure 3. *MSE loss of the model predicting the daily positivity ratio over the 100 epochs of training*

B. Scoring

Table 2 reports the prediction accuracy metrics for new cases for the period Aug 30, 2021 - Sep 26, 2021. Please note, that these are subject to slight variability as the initial model weights used are randomized.

Table 2. *Root Mean Squared Error & Mean Absolute Error of the proposed model for positivity ratio*

Root Mean Squared Error	Mean Absolute Error
2.45×10^{-3}	1.53×10^{-3}

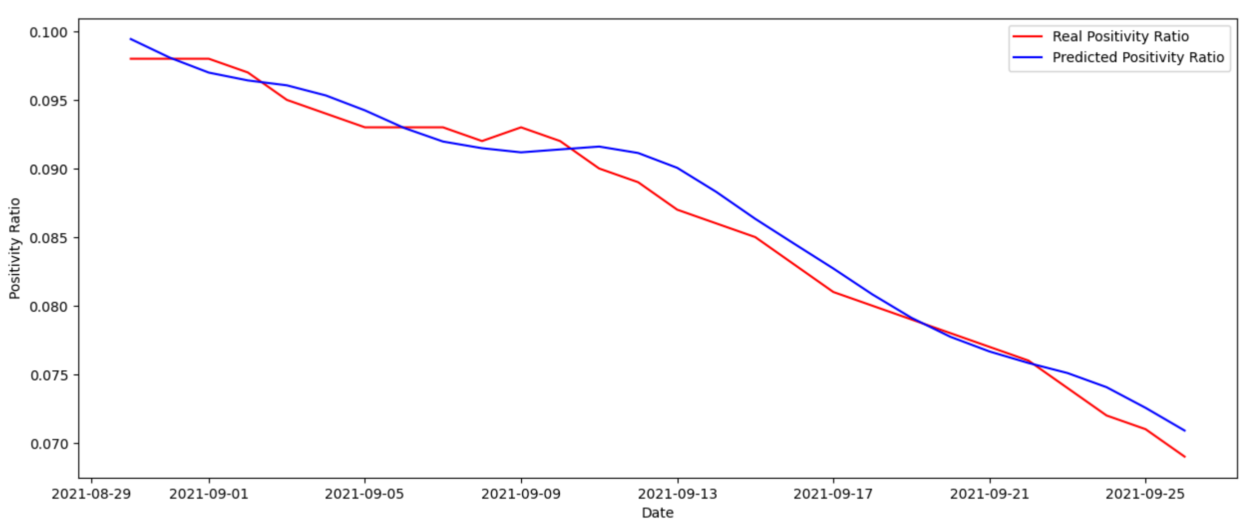

Figure 4. Model Prediction and Actual Positivity Ratio