**Literature Review**

Our team reviewed several articles regarding forecasting models to project the spread of the covid-19 pandemic. From the outset in March 2020, researchers, data analysts, and machine learning engineers applied various time series models from Simple Exponential Smoothing to ARIMA to Deep Learning Neural Networks to Facebook’s Prophet all in an attempt to alert public health officials about the spread of the virus and predict a possible end date. As stated in the article, “COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population” by Vasilis Papastefanopoulos , Pantelis Linardatos and Sotiris Kotsiantis, the authors state: “Being able to accurately forecast when the outbreak will hit its peak would significantly diminish the impact of the disease, as it would allow governments to alter their policy accordingly and plan ahead for the preventive steps needed such as public health messaging, raising awareness of citizens and increasing the capacity of the health system.”[[1]](#footnote-1)

Our review, which also included covid-19 time series projects on Kaggle, discovered the following similarities across all of the materials:

*Data Collection*

Data from Johns Hopkins University (JHU) and the World Health Organization (WHO) were the most common source of information for these articles. The data covered the time period from the early days of the pandemic to the summer of 2020. With the benefit of hindsight, our model will be able to expand on the time frame of data collection as well as have access to more complete data. It is an established fact that there were major problems with data collection in the early days of the pandemic. The covid tracking project at The Atlantic magazine created a scoring system for how states reported their data.[[2]](#footnote-2) Since March 7, 2021, The Atlantic covid tracking project has stopped tracking data due to the fact that the CDC has improved its data collection and reporting.

*Models*

Several different mathematical and time series models were used to forecast the infection rates. Ping-En Lu Yi-Cheng Chen and Cheng-Shang Chang’s paper, “A Time-Dependent SIR Model for Covid-19 with Undetectable Infected Persons”, used the time-dependent model, susceptible-infected-removed (“SIR”) that tracks transmission and recovery rates. Ian Cooper,, Argha Mondal, and Chris G. Antonopoulos also applied the SIR model in their paper, “A SIR model assumption for the spread of COVID-19 in different communities “. They “augmented the classic SIR model with the ability to accommodate surges in the number of susceptible individuals, supplemented by recorded data from China, South Korea, India, Australia, USA, Italy and the state of Texas in the USA to provide insights into the spread of COVID-19 in communities. In all cases, the model predictions could be fitted to the published data reasonably well, with some fits better than others.”

The most common time series approach is exemplified in the work by Vasilis Papastefanopoulos, Pantelis Linardatos, and Sotiris Kotsiantis. They used six different time series approaches, Auto-Regressive Integrated Moving Average (“ARIMA”), the Holt–Winters additive model (“HWAAS”), TBAT (an acronym for the four fundamental components of the method: Trigonometric seasonal formulation, Box–Cox transformation, ARMA errors and trend component), Facebook’s Prophet, DeepAR (a forecasting method based on autoregressive recurrent neural networks for probabilistic forecasting), and N-Beats (Neural Basis Expansion Analysis for Interpretable Time Series Forecasting). The models were applied across 10 nations. Using RMSE to assess the models, they found “traditional statistical methods such as such ARIMA and TBAT overall prevail over deep learning counterparts such as DeepAR and N-BEATS—an outcome which, due to the lack of large amounts of data.”

Our models will improve on the work that we have reviewed so far given that the quality of data collection has improved, and we have more information to tune the various time series models to make our predictions.

**Key Words**

*Pandemic, Covid, Time Series, MAPE, RMSE, ARIMA, Prophet, Cases, Predictions, SARIMA, Forecasting, Comparison, Long short-term Memory (LSTM), Recurrent Neural Network (RNN), Bidirectional LSTM (BiLSTM), Gated Recurrent Units (GRUs), and Variational AutoEncoder (VAE), Support Vector Machines (SVM*

1. Papastefanopoulos, Vasilis, Linardatos, Pantelis, and Kotsiantis, Sotiris, “COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population” Applied Sciences. 10. 3880. 10.3390/app10113880 (May 2020) [↑](#footnote-ref-1)
2. https://covidtracking.com/analysis-updates/weve-launched-a-new-state-grading-system [↑](#footnote-ref-2)