Analysis and Predictive Modeling of COVID-19 and its Effects

Aman Aryan - 20111009
Deeksha Arora - 20111017
Shruti Sharma - 20111061
Tamal Deep Maity - 20111068

Indian Institute of Technology, Kanpur

December 14, 2020

Introduction

The novel Coronavirus-2019 outbreak which was reported by World Health Organization (WHO) on 21st January, 2020 has spread around the globe since months now and caused serious health, social and financial issues.

The epicentre was found to be in Wuhan, China. Regions across the world applied restrictive measures such as closing public places, schools, universities, quarantining individuals etc. The global economy also went downhill.

```
Total confirmed cases till 12/8/20 are 68225313
Total deaths till 12/8/20 are 1556822
Therefore, mortality rate is 2.281883265233243
```

Problem Statement

To contain the spread of the virus, it is extremely important to deploy technology for its future trends.

Tracking the active corona cases as they crop up helps governments and common people be cautious about their future plans. They may then decide and form policies for better control over the virus. It will lead to less havoc and lower fatality rate.

In this project, we aim to develop a probabilistic predictor for confirmed coronavirus cases in future 1 to 30 day window. Our objective is to correlate this prediction with the GDP projection in that particular region.

Literature

The SIR family of models are among the most popular ones to learn the dynamics of COVID-19. The SIRD model is an extension of the SIR model, and is used to demonstrate the differences of infectious rate at different countries.
There are models available based on Markov chain Monte Carlo methods and other kinds of Bayes filters like Particle filters and Kalman filters.

Dataset

- All reported experiments are based on online COVID-19 data repository maintained by John Hopkins University - `https://github.com/CSSEGISandData/COVID-19` - for time-period of 22 January 2020 to 8 December 2020.
- Dataset for GDP Projection: [1]
  - Gross Domestic Product(Real Index) - IMF Data
  - Total Population (2019) (Worldbank)
  - Unemployment (Worldbank)
  - Unemployment [FOR INDIA]
  - Mobile cellular subscriptions (Worldbank)
  - Population age 65 and above data (Worldbank)

[1] links for dataset on last slide

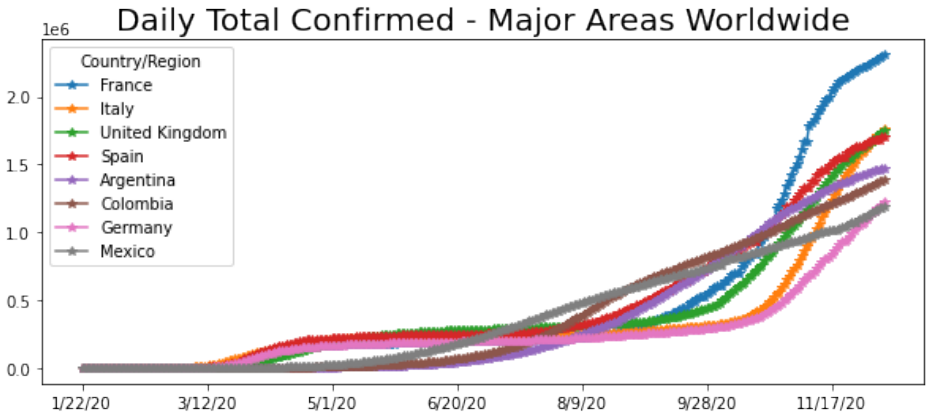
Methodology

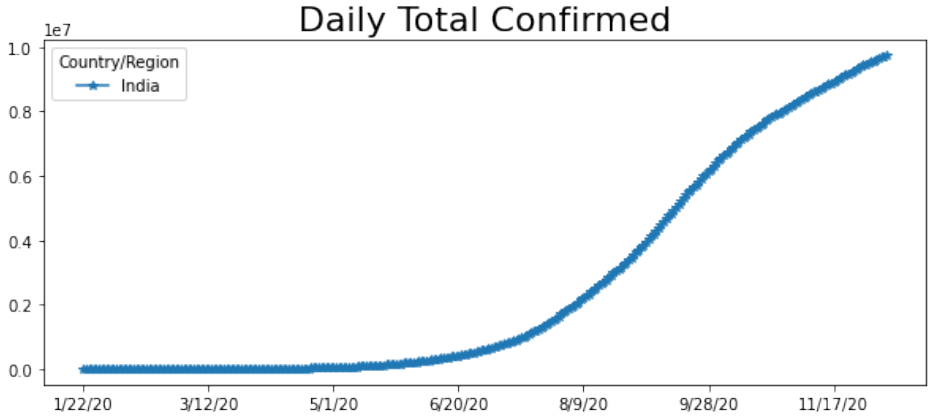
- The dataset was filtered and structured.
- Time-series analysis of data to decipher trends.
- An adaptive online Kalman filter provides us a very good one-day prediction for each region.
- It was extended for a window of 30 days which showed how Kalman filter behaves undesirably for long-term predictions.
- The results of this predictor are used in projecting the GDP of a region.

Data Preprocessing

- Read the data containing daily total cases of confirmed, death and recovered.
- Fix region names.
- Consider regions for which reliable population data is available.
- Missing values are handled using interpolation.
- Data was scaled quarter-wise from the available annual data using interpolation.

Visualisations





We can see a steep growth of active cases in India from July as the lockdown was lifted.

Kalman Filters

Kalman filter (KF) is a widely used method for tracking and navigation, and for filtering and prediction of econometric time series data. Kalman filter is a recursive algorithm that uses time-series measurement over time, containing statistical noise and produce estimations of unknown variables.

Each day the algorithm needs to be updated with new observation, after the estimation is done it can generate predictions for the next day.

No overfitting or bias as it is an online algorithm.

Error function : RMSE and MAE

Features

- Confirmed cases
- One day change
- Three day change
- Five day change
- Seven day change
- One day change rate (100*(One day change) / (Cases 1 day ago))
- Three day change rate
- Five day change rate
- Seven day change rate
- Infected rate
- Kalman Prediction

Results

Training Date - 22 January, 2020 to 7 December, 2020
Testing Date - 8 December 2020 to 9 December, 2020

	0	confirmed	date	Country/Region
9768315		0	2020-12-09	India

Figure: 1-Day Prediction for Confirmed Cases, India

Actual cases on 2020-12-09 in India = 9771292

Results

date	India
2020-12-09	9735850.0
2020-12-10	9735850.0
2020-12-11	9735850.0
2020-12-12	9989330.0
2020-12-13	10193229.0
2020-12-14	10455492.0
2020-12-15	10455492.0
2020-12-16	10455492.0
2020-12-17	10455492.0
2020-12-18	10773676.0
2020-12-19	10999149.0
2020-12-20	10999149.0
2020-12-21	10999149.0
2020-12-22	10999149.0
2020-12-23	11610589.0
2020-12-24	11906450.0
2020-12-25	12420226.0
2020-12-26	12948398.0
2020-12-27	12948398.0

Figure: 30-day Prediction Table for Confirmed Cases, India

Results

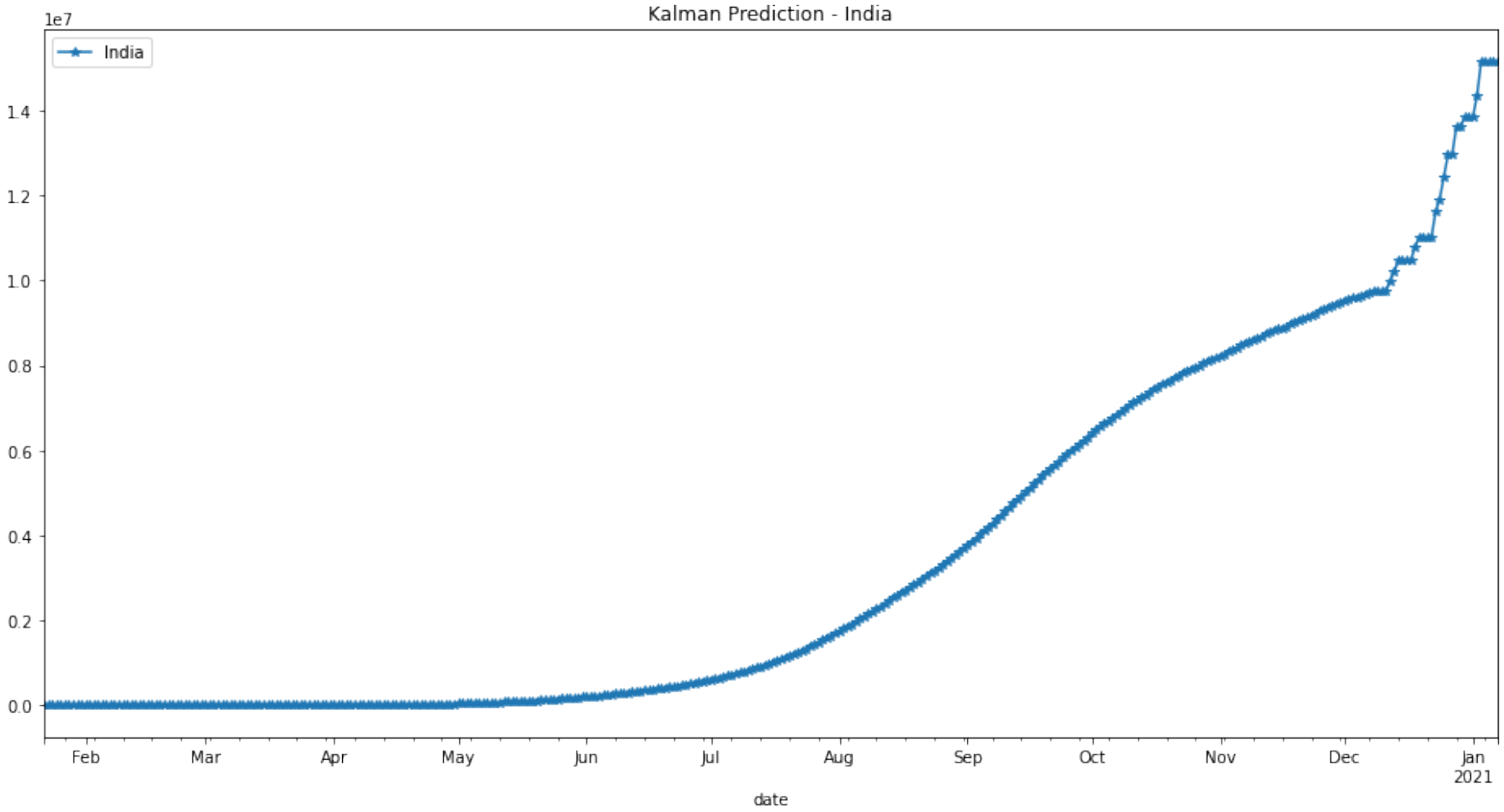


Figure: Prediction Curve for India (Long-term)

Results

	Country/Region	rmse	mae
33	China	73.924522	8.080997

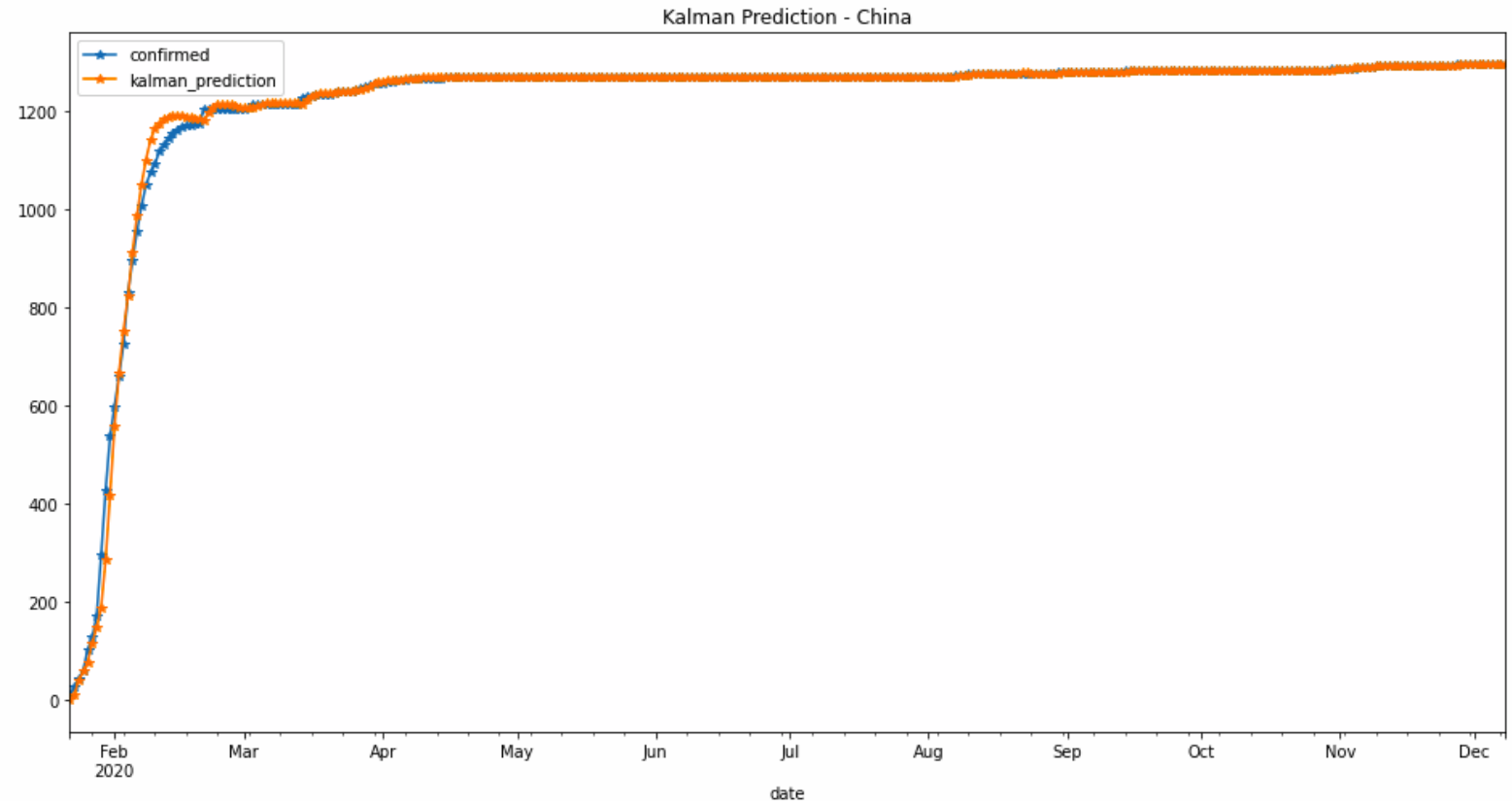


Figure: Prediction Curve for China (Short-term)

Neural Net

Architecture

- An input layer, 3 hidden layers, An output layer
- Activation function - LeakyRELU
- Optimizer - Adam
- Learning Rate - 0.001
- Cost function - MSE

Results

Date	Confirmed(Predicted)	Confirmed(Actual)	Deaths(Predicted)	Deaths(Actual)
04 Dec	69861000	68225587	1498820	1556834
05 Dec	70584000	None	1509090	None
06 Dec	71316000	None	1519450	None
07 Dec	72055000	None	1529890	None
08 Dec	72802000	None	1540370	None
09 Dec	73556000	None	1550940	None
10 Dec	74319000	None	1561580	None
11 Dec	75089000	None	1572300	None
12 Dec	75891000	None	1583080	None
13 Dec	76727000	None	1593940	None
14 Dec	77606000	None	1604880	None

Figure: Prediction Table for Global Confirmed and Death Cases

Results

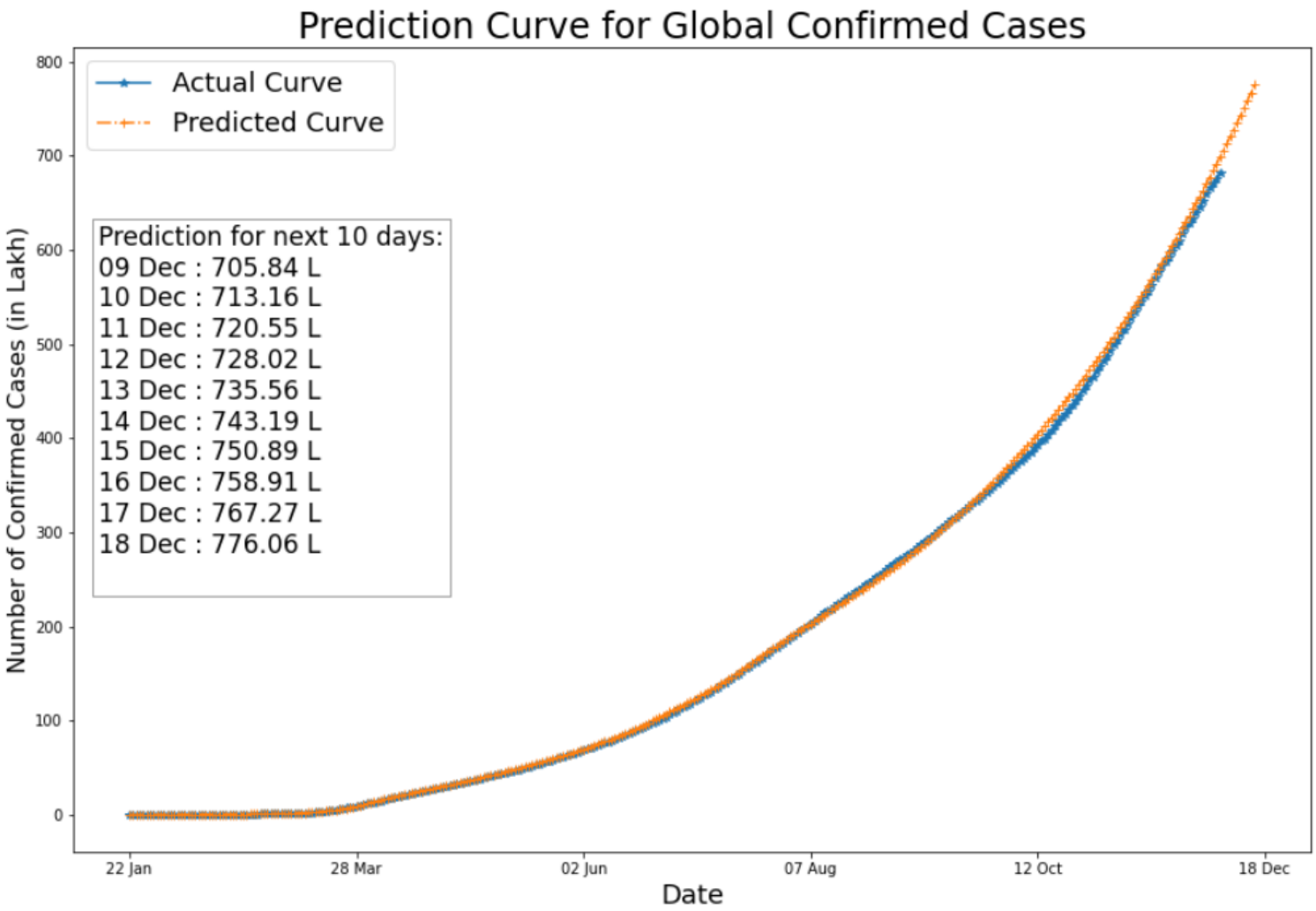


Figure: Prediction Curve for Global Confirmed Cases

Results

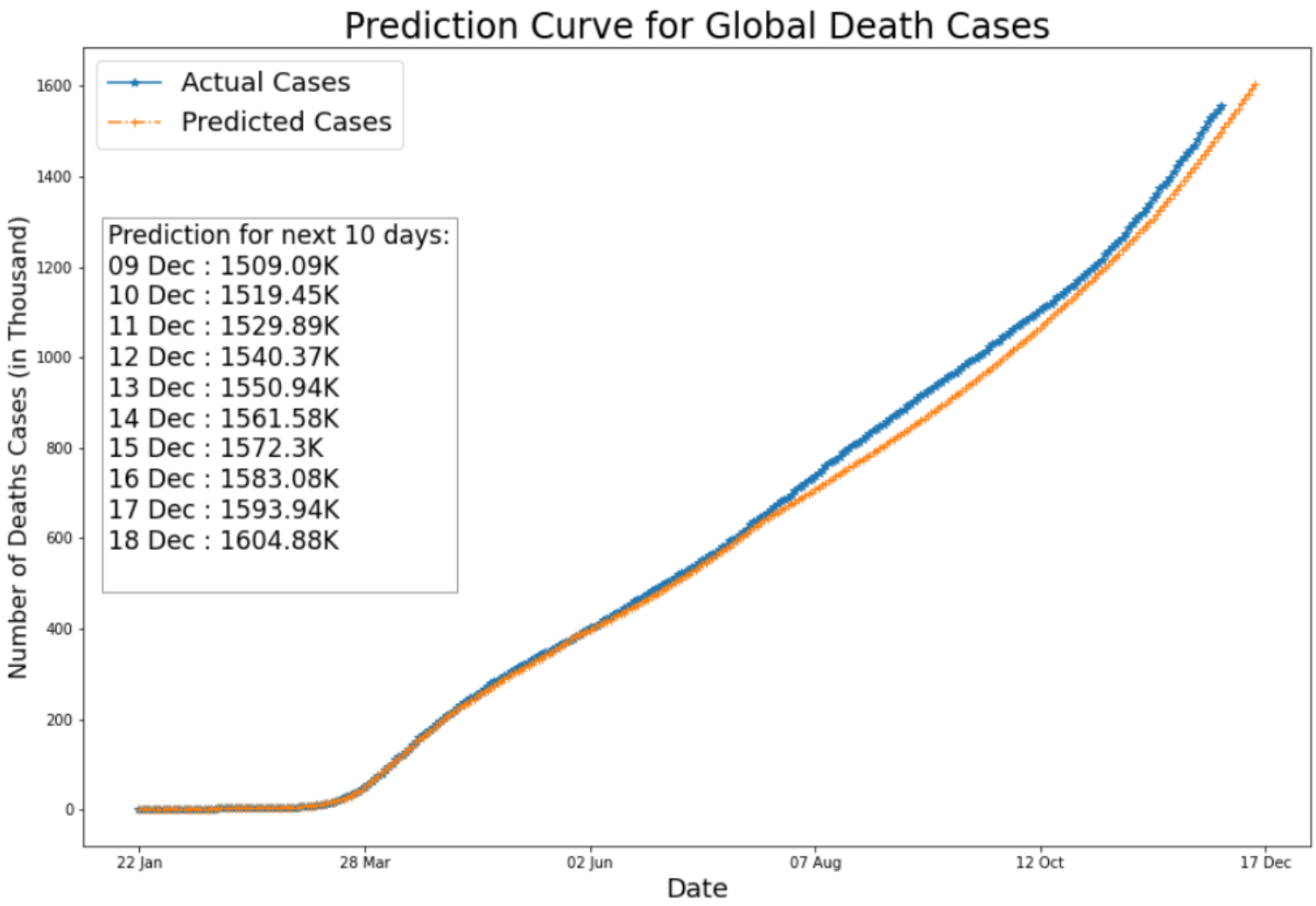


Figure: Prediction Curve for Global Death Cases

GDP Predictor

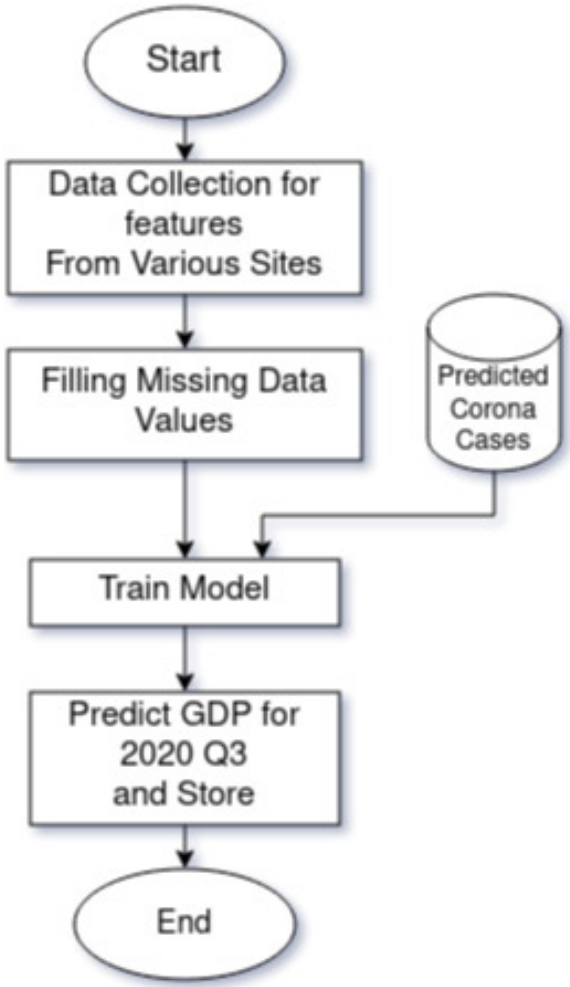
Features used for training the model

- Total Population
- Unemployment Ratio
- Mobile cellular subscriptions
- Population above 65 years of age
- Predicted COVID-19 cases (for 2020 Quarter 3)

Random Forest Regressor

- Ensemble learning method for classification
- Can be used for regression too
- Constructs several DTs at training time
- Outputs mean of individual trees
- Hyperparameters: Estimators $= 4$, Criterion : MSE

Flowchart for GDP Prediction



Results

Year/Quarter	2020Q1	2020Q2	2020Q3_Predicted
Australia	3.342654	-5.848295	0.312897
Brazil	-1.500000	-9.600000	-3.016573
Canada	1.432935	-11.300000	-3.917316
United States	2.006274	-5.000000	0.105028
United Kingdom	0.645599	-13.549770	-4.747919
France	-3.200245	-15.594570	-6.518146
Spain	-2.642289	-20.632573	-7.573589
India	-22.569823	-7.000000	1.366490
Singapore	-0.800000	-13.200000	-3.519417
Thailand	-1.328246	-14.488241	-4.469328
China	-6.800000	3.200000	6.084559
Greece	-2.742718	-17.112160	-7.683863

Figure: Prediction of percent change in GDP of Countries

Discussion

- Population and daily total confirmed cases are used to calculate the infected rate of each area (in percentage).
- Kalman prediction and time-dependent variables like day-wise changes are correlated to the target i.e. confirmed cases.
- The 1-day Kalman prediction is very accurate and powerful while a longer period prediction is more challenging but provides a future trend.
- Large mean average error for longer predictions.
- Historical data has less effect on prediction as compared to the previous day data.
- GDP projection is somewhat consistent with actual projections for countries that provided comprehensive data about the influencers.

Future Work

- Take testing scale into account.
- Correlate the predictions with other factors like healthcare, age groups etc.
- Continuous evaluation of GDP instead of quaterly evaluation.

External Links

- Gross Domestic Product(Real Index) - https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b
- Total Population (2019) - api.worldbank.org/v2/en/indicator/SP.POP.TOTL?downloadformat=csv
- Unemployment - api.worldbank.org/v2/en/indicator/SL.UEM.TOTL.ZS?downloadformat=csv
- Unemployment [FOR INDIA] - https://www.macrotrends.net/countries/IND/india/unemployment-rate
- Mobile cellular subscriptions - api.worldbank.org/v2/en/indicator/IT.CEL.SETS.P2?downloadformat=csv
- Population age 65 and above data - api.worldbank.org/v2/en/indicator/SP.POP.65UP.TO.ZS?downloadformat=csv

References

https://link.springer.com/article/10.1007/s10489-020-01948-1

THANK YOU