Disease Outbreaks

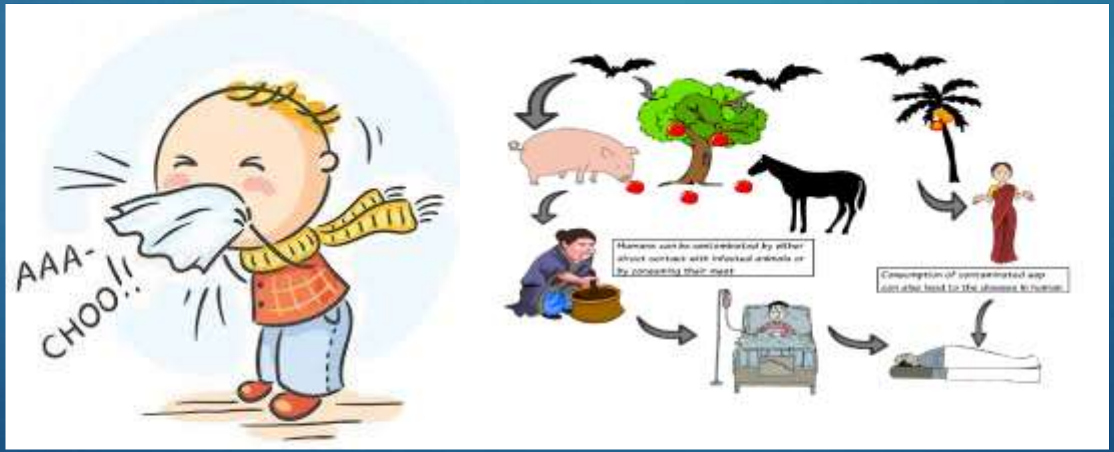
NAVANEETH M
NIROSHA BUGATHA
KUNAL NIRANJAN DESAI
ARJUN MAHADEVAN

Motivation

- What is a Disease outbreak?
  - A disease outbreak is an epidemic that occurs suddenly and within a relatively confined geographic area. The occurrence of disease outbreak depends on the mode of transmission , source of disease etc.,
- COVID-19 pandemic that started in January 2020.
- To gain insights into fatalities and factors affecting outbreaks between 1996 to 2020.
- To create awareness about different diseases and its epidemiological profile like mode of transmission, symptoms, pathogens etc.,

Questions

- What were the most fatal outbreaks experienced around the world in each country ?
- How does COVID-19 compare to past outbreaks ?
  - Case Fatality, Spread, Transmission Medium, Incubation Period, etc.
- Do some countries have more experience handling outbreaks ?
- Can we model the spread of a disease ?



Methodology

Phase I

Data Collection
- Web Scraping
- NLP-Entity Extraction from News Articles

ETL
- Dataset creation, Identify Features

Visualization
- Visualization & Information Dashboard

Phase II

Model for Fatality
- Cluster Outbreak based on diseases and country
- Case Fatality Ratio

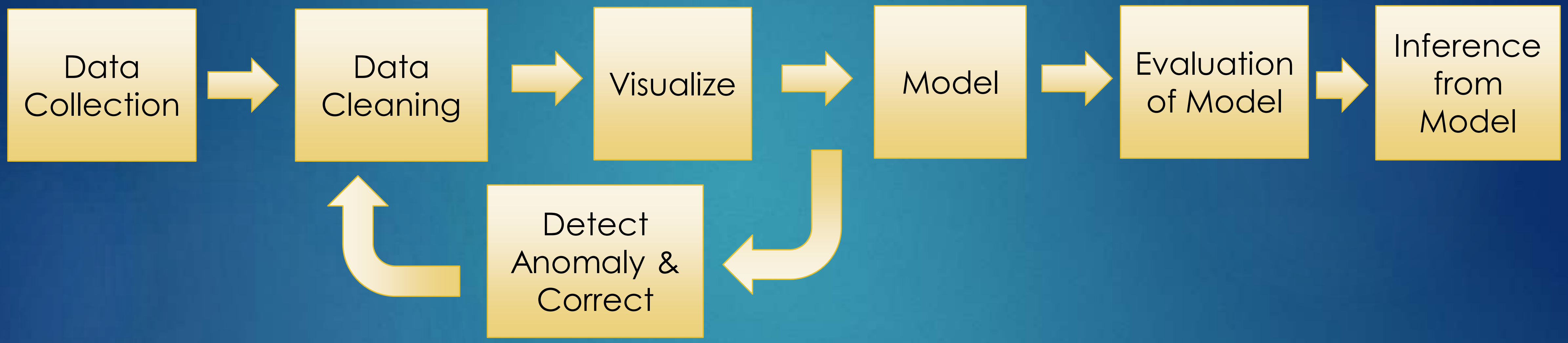
Model for Spread
- Compartmental SIR disease model using infection rate and recovery rate

Inference
- Use model to infer spread and compare fatalities

Datasets/Tools

- Data Source:
  - World Health Organization
  - CDC Centers for Disease Control and Prevention — CDC 24/7: Saving Lives. Protecting People™
  - Johns Hopkins University & Medicine | Coronavirus Resource Center
- Tools:
  - Power BI
  - scikit learn
  - python
  - Beautiful Soup
  - spaCy

Data-science pipeline



Results & Evaluation

Past Outbreaks & Fatalities

Most Fatal Outbreaks around the world between 1996 to 2020

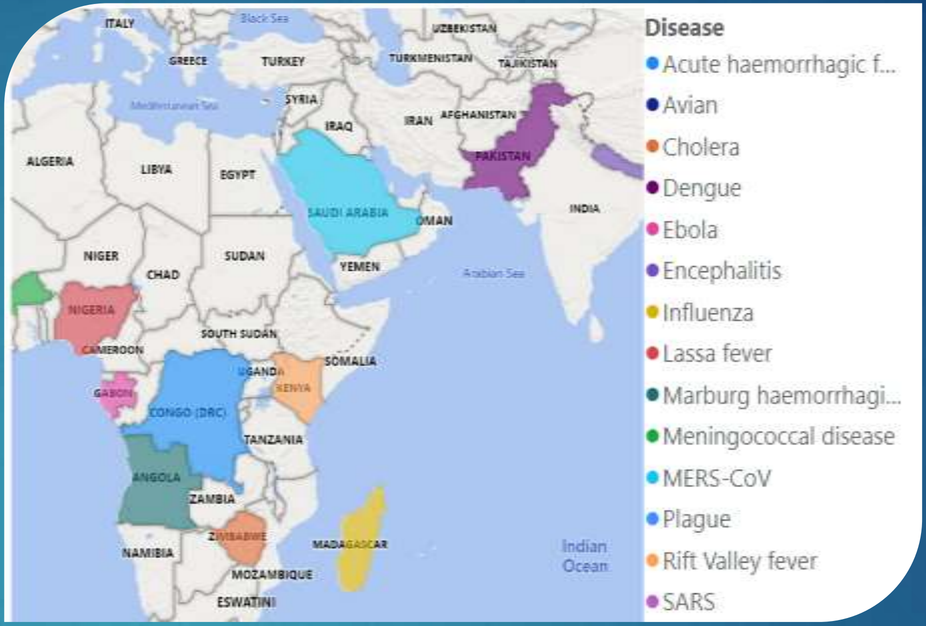
Year	Country	Disease	CFR	Death
1996	Nigeria	Cholera	92.8	487
1996	Gabon	Ebola	66.7	42
1996	Benin	Yellow fever	47.7	41
1996	India	Dengue	38.1	354
1996	Sierra Leone	Lassa fever	5.7	14

Page 1

Disease_Outbreak_Dashboard_modifications_AM_KD | Data updated 4/7/20

Outbreaks with Highest Case Fatality Ratios

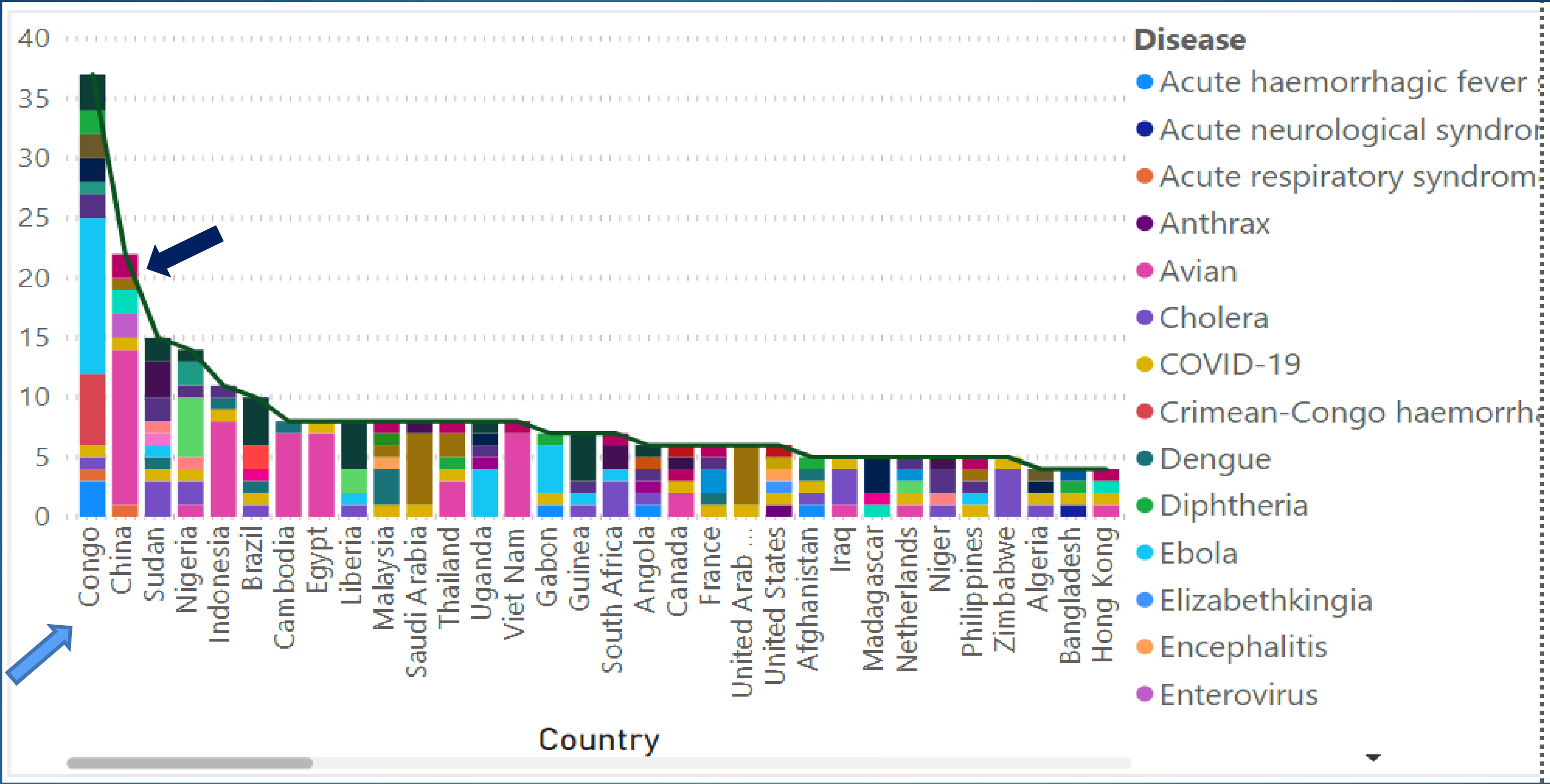
Disease	Country	Year	Death	CFR
Encephalitis	Nepal	1997	223	90.28
Ebola	Congo	2003	128	89.51
Marburg haemorrhagic fever	Angola	2005	357	84.40
Influenza	Madagascar	2002	671	83.04
Avian	Indonesia	2008	113	81.75
Meningococcal disease	Burkina Faso	2002	447	76.15
Acute haemorrhagic fever syndrome	Congo	2003	44	72.13
Viral haemorrhagic fever	Congo	1999	63	67.74
Ebola	Gabon	1996	42	66.67
Meningococcal disease	Burkina Faso	2003	542	55.31
Lassa fever	Nigeria	2016	149	54.58
Plague	Madagascar	2017	202	41.94
MERS-CoV	Saudi Arabia	2014	347	36.03
Yellow fever	Brazil	2019	483	35.10
Rift Valley fever	Kenya	2007	121	29.44
West Nile	United States	2002	211	21.91
Typhoid fever	Congo	2004	134	19.25
SARS	Hong Kong SAR, China	2003	298	16.98
SARS	Taiwan	2003	735	8.94
Dengue	Pakistan	2019	3075	6.53
Cholera	Zimbabwe	2009	731	0.92



COVID-19

Country	CFR
Italy	12.72
UK	12.47
France	10.78
Spain	10.30
Iran	6.24
China	4.02
US	3.95
Germany	2.35

Countries that Experienced Most Outbreaks



- For some countries, the COVID-19 is the first disease outbreak, while some countries have had a disease outbreak almost every year
- Congo has experienced highly fatal outbreaks
- China has had a high number of low fatality Influenza type outbreaks.

Time Dependent SIR Model For COVID-19

- SIR model is a compartmental model that describes the dynamics of an infectious disease. The model divides the population n into compartments S (Susceptible), I (Infectious) and R (Recovered).

$$S(t) + X(t) + R(t) = n$$

- SIR model describes the number of people in each of the above compartments using differential equations.

$$\frac{dS(t)}{dt} = \frac{-\beta S(t)X(t)}{n}, \qquad \frac{dX(t)}{dt} = \frac{\beta(t)S(t)X(t)}{n} - \gamma X(t), \qquad \frac{dR(t)}{dt} = \gamma X(t)$$

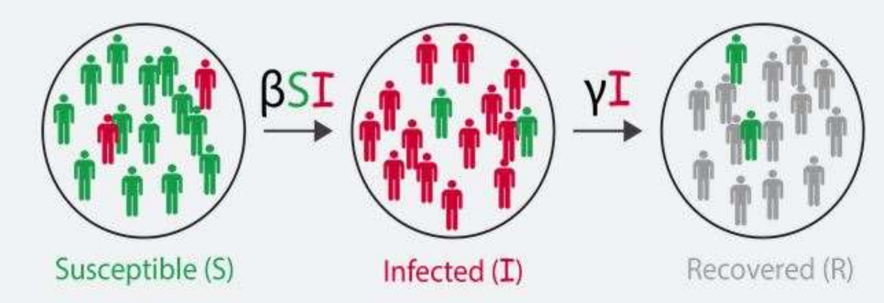
- Transmission Rate (β) is a parameter that determines the disease transmission through exposure and contact.

$$\beta(t) = \frac{[X(t+1) - X(t)] + [R(t+1) - R(t)]}{X(t)}$$

- Recovery Rate (γ) is a parameter that expresses the rate of disease recovery.

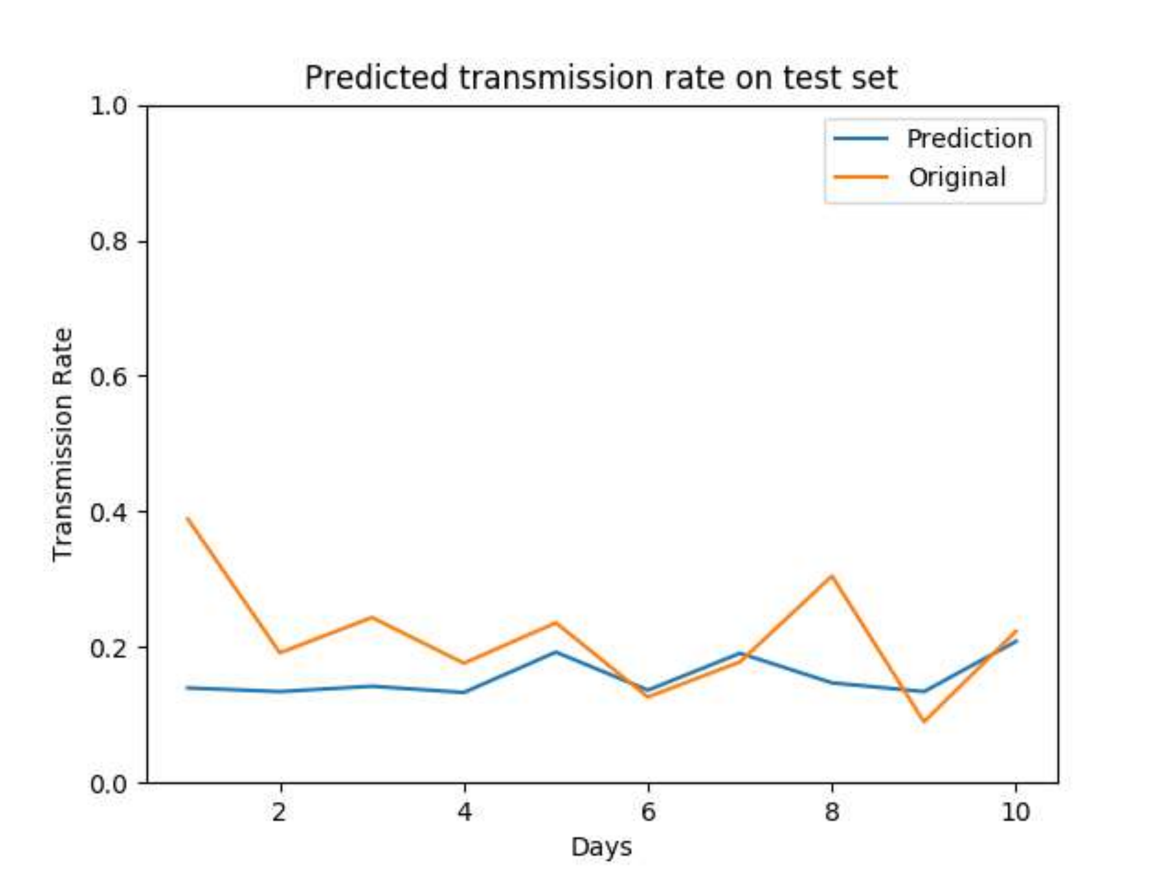
$$\gamma(t) = \frac{R(t+1) - R(t)}{X(t)}$$

- We used data from first 45 days to learn the parameters β and γ.
- $\hat{\beta} = a_1\beta(t-1) + a_2\beta(t-2) + \ldots + a_J\beta(t-J) + a_0 = \sum_{j=1}^{J} a_j \beta(t-j) + a_0$
- $\hat{\gamma} = b_1\gamma(t-1) + b_2\gamma(t-2) + \ldots + b_K\gamma(t-K) + b_0 = \sum_{k=1}^{K} b_k\gamma(t-k) + b_0$

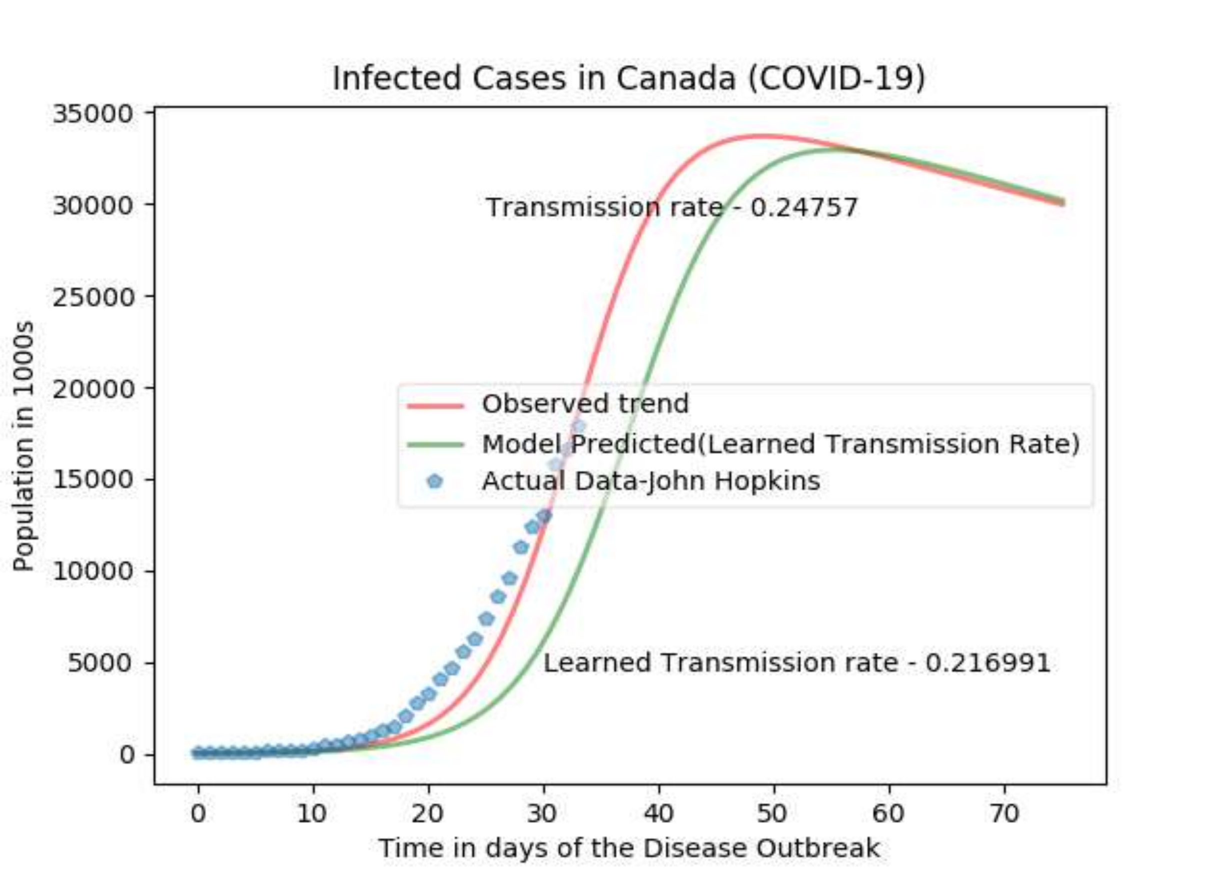


Insights From SIR Model - Canada

Predicted Transmission vs Original Transmission Rate on Test set

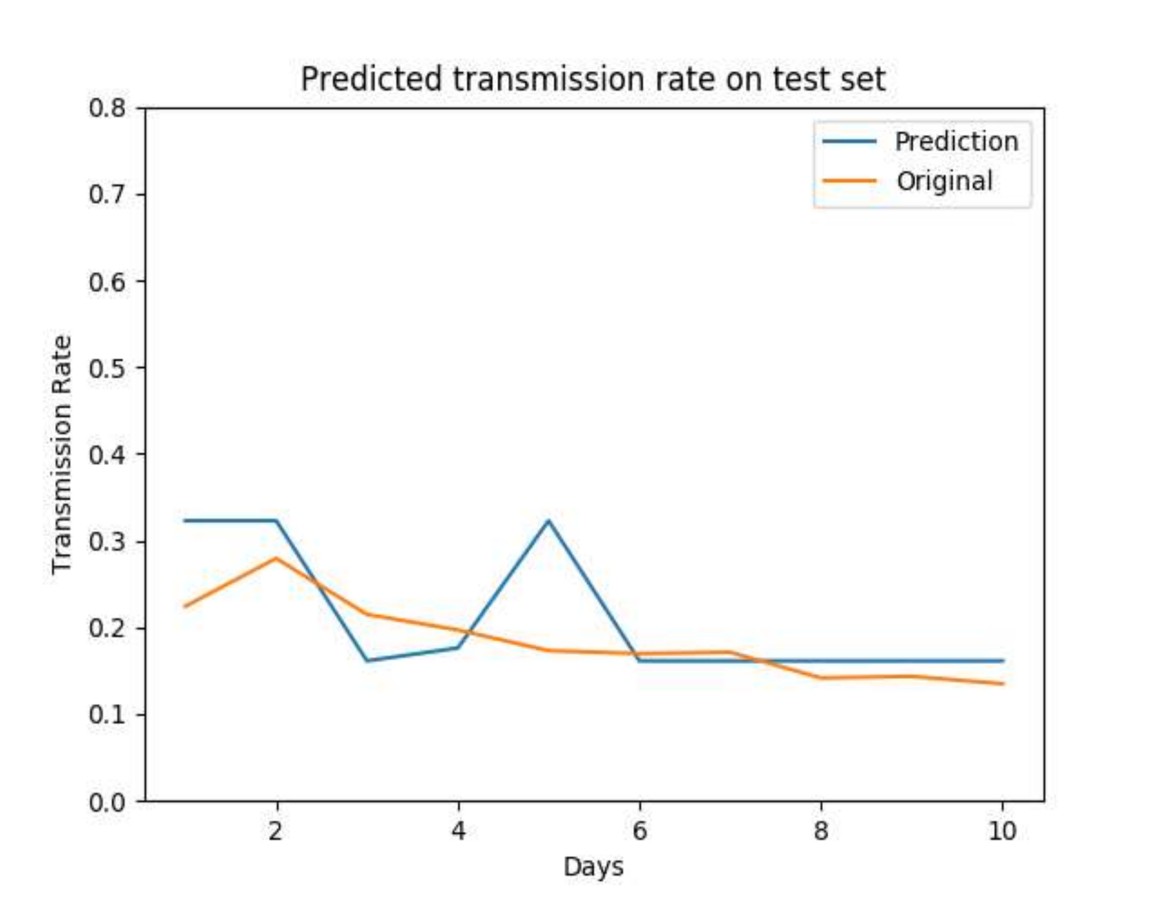


Variation in the Observed Trend and Predicted Trend of Confirmed Cases

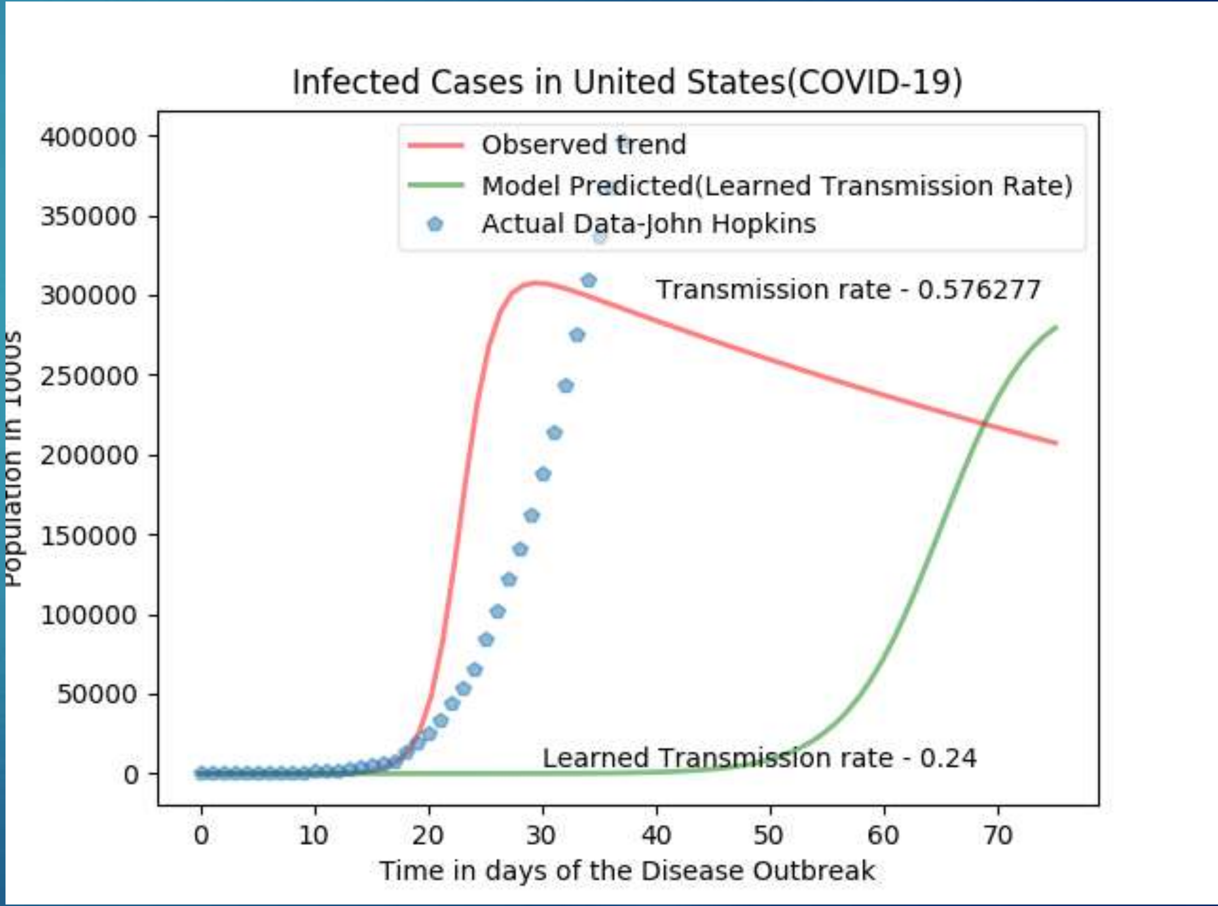


Insights From SIR Model - USA

Predicted Transmission vs Original Transmission Rate on Test set



Variation in the Observed Trend and Predicted Trend of Confirmed Cases



Why result makes sense?

- Our project captures the incompleteness with respect to reporting of data during an epidemic
- Government officials and decision makers rely on the reported data available which do not portray the true picture during an epidemic i.e., the true spread could be much higher

Data product

- Publically available data set on outbreaks other than covid-19
- Dashboard of all outbreaks from 1996 to 2020

Key Learning

- Formulating Data from scratch from different websites when no data was available.
- Identifying tools to extract features and entity in best possible and efficient manner
- Ability to deliver complex information through Dashboard
- Can we still use traditional models for disease analysis? Identified epidemiological models
- Formulated probabilistic models and regression data
- In short span of time making our project relatable and present results for current COVID-19 situation as well

Future Work

- Live updating of dashboard based on availability of data for COVID-19
- Update model parameters for COVID-19 as and when more data becomes available from all over world
- Investigate further to improve model to understand spread of epidemic

Thank you

☺