Results Presentation

Team: PandeML

Lukas Codispoti, Micha Weber, Costanza Maria Improta, Philip Leindecker, Ole Müller

Outline

1) Chosen Approaches
    a) Parametric Curve Fitting
    b) Linear Regression
    c) SIR
    d) LSTM
2) Comparison of Models
3) Conclusion

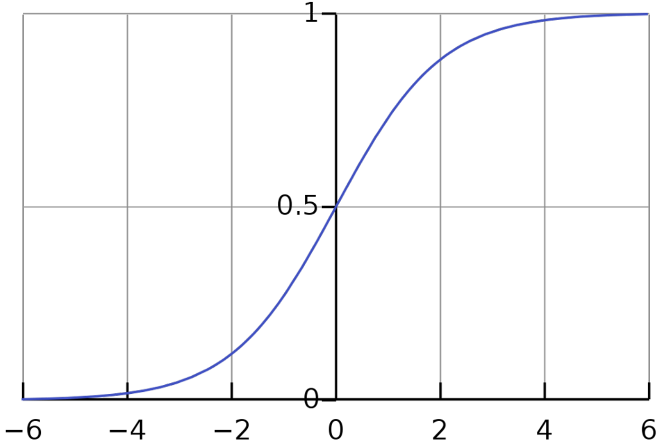
Parametric Curve Fitting

Idea: Number of infected/deceased/recovered people will follow a **Logistic Curve.**
→ Fit data to a such a curve using scipy's `optimize.curve_fit`

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Parameters:

L :=	Maximal value
k :=	Logistic growth rate
x_0:=	Midpoint



Credit: https://en.wikipedia.org/wiki/Logistic_function

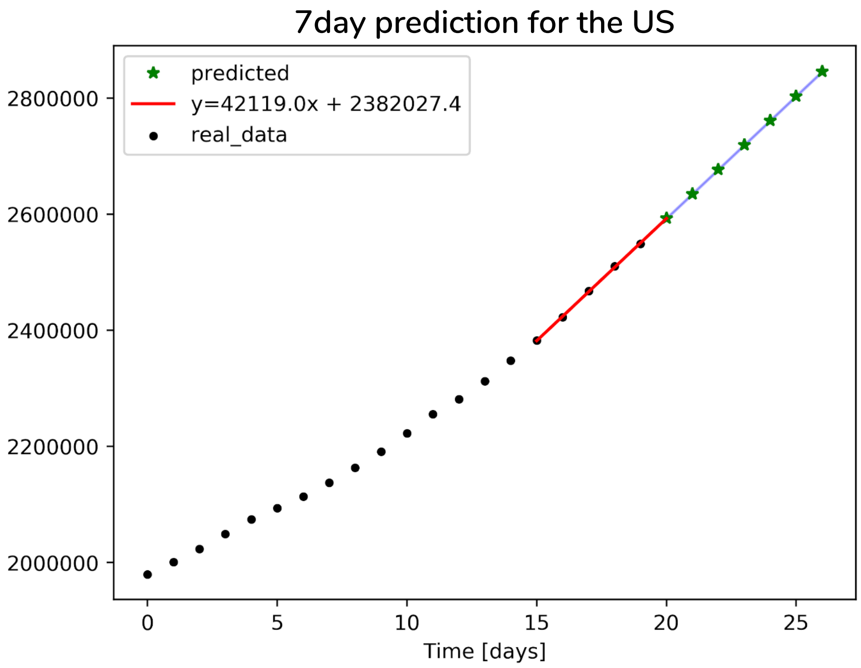
As proposed in https://medium.com/analytics-vidhya/how-to-predict-when-the-covid-19-pandemic-will-stop-in-your-country-with-python-d6fbb2425a9f

Linear Regression

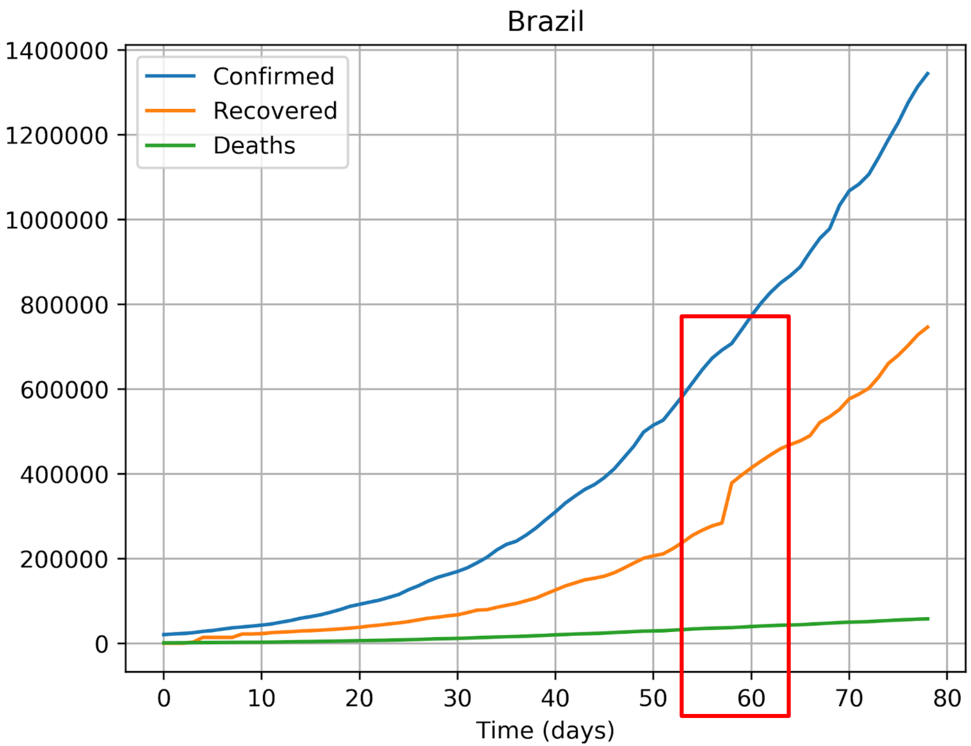
- **Motivation**: Simpler models tend to perform better for short time predictions
- **Model**: Ordinary least squares linear regression:

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 = \min_{\boldsymbol{w}\in\mathbb{R}^p} \sum_{i=1}^{n} (\boldsymbol{w}^\top \boldsymbol{x}^{(i)} - y_i)^2.$$

- **Parameters**: LinReg of last 5 days for both 2day and 7day prediction
- Usage of **scipy.stats.linregress**, to determine: slope, intercept and std_err
- **95%-Confidence-Interval** via: 1.96*std_err
- **Pros**: Mathematically straightforward, Transparent
- **Cons**: Not capable to predict complex relationships (in opposite e.g. SIR)

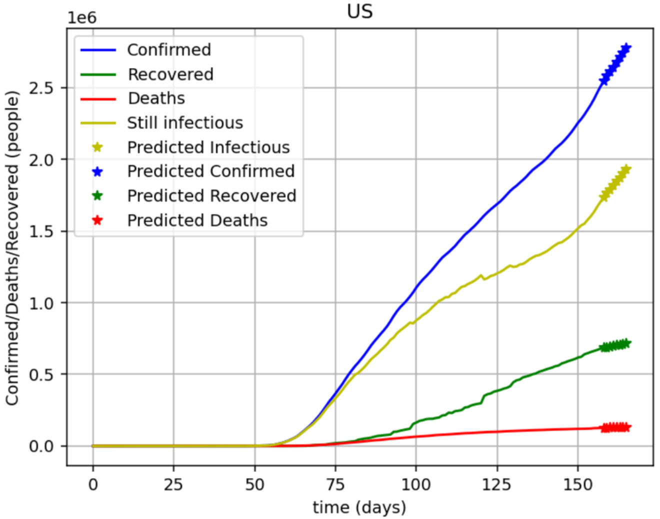


Linear Regression



SIRD

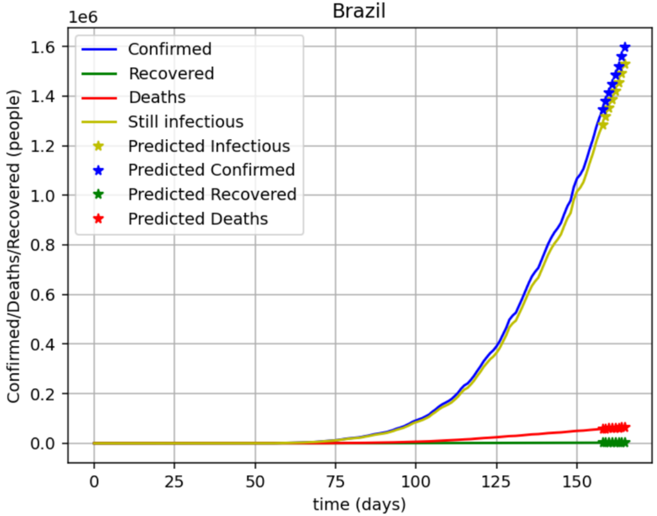
$$\frac{dS}{dt} = -\beta \cdot I \cdot \frac{S}{N}$$
$$\frac{dI}{dt} = -(1-\alpha) \cdot \gamma \cdot I - \alpha \cdot \rho \cdot I$$
$$\frac{dR}{dt} = (1-\alpha) \cdot \gamma \cdot I$$
$$\frac{dD}{dt} = \alpha \cdot \rho \cdot I$$



➢ β = number of persons an infected person infects per day

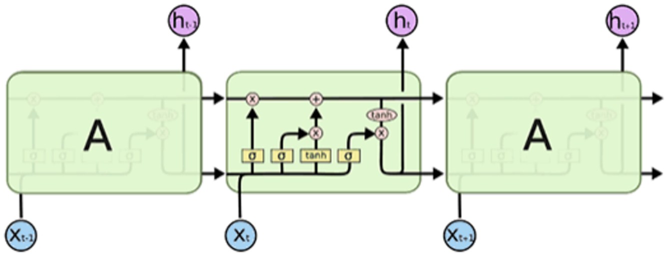
➢ days = number of days an infected person is infectious

➢ R = beta × days = total number of persons an infected person infects



Problem:

- SIR-model tries to describe "real" epidemic outbreak, but prediction is made for incomplete data (mainly missing recovered data) → parameters need to be calculated from given data → leads to unreal parameters
- Example:
  - Calculated parameters for prediction in US:
    R = 7.13, β = 0.018, days = 396.8



LSTM-Recursive NN

LongShortTermMemory

<u>Idea</u>: Use COVID19 data from all countries world wide to learn the behavior
(11000 training Samples, 2000 Samples for validation)

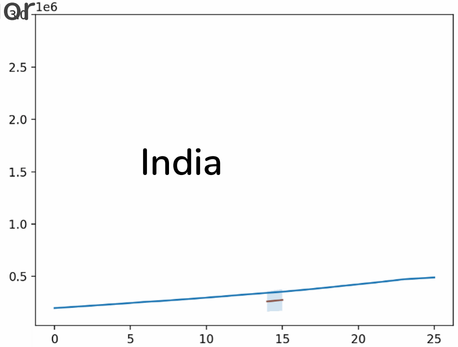
<u>Strategy</u>: *Multiple Parallel Input and Multi-Step Output* for 100 epochs

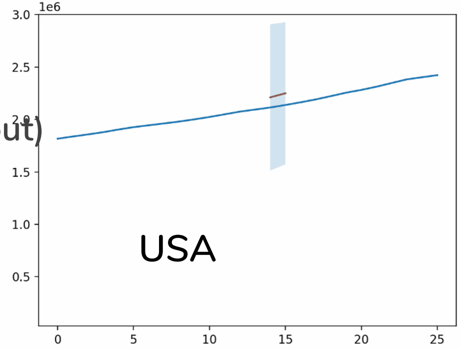
<u>Input</u>: 7 Days of data (Cases, Deaths, Recovered)

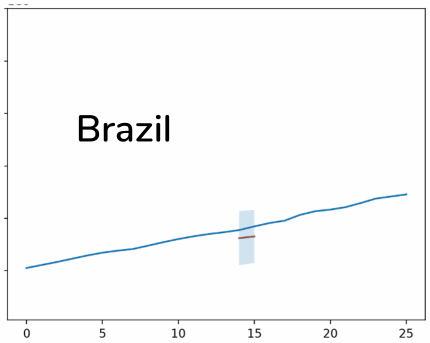
<u>Output</u>: 2 or 7 Days of Data (Cases, Deaths, Recovered)

<u>Architechture:</u>

200 Node LSTM
150 Node LSTM
100 Node LSTM
50 Node LSTM
3 Node Dense (output)

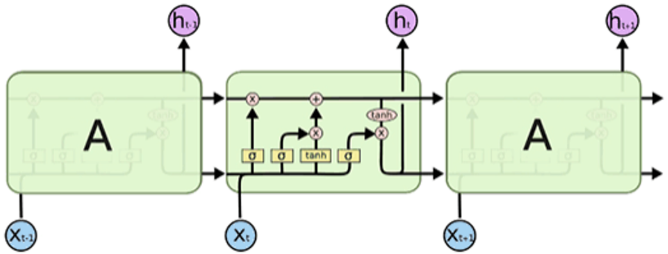






LSTM-Recursive NN

LongShortTermMemory



Idea: Use COVID19 data from all countries world wide to learn the behavior (11000 training Samples, 2000 Samples for validation)

Strategy: *Multiple Parallel Input and Multi-Step Output* for 100 epochs

Input: 7 Days of data (Cases, Deaths, Recovered)

Output: 2 or 7 Days of Data (Cases, Deaths, Recovered)

Architechture:

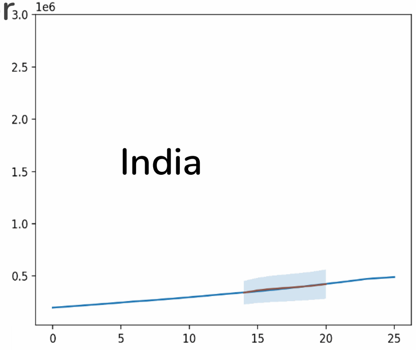
200 Node LSTM

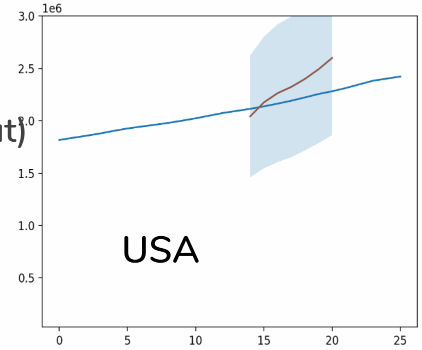
150 Node LSTM

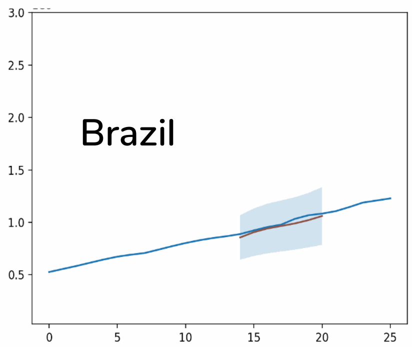
100 Node LSTM

50 Node LSTM

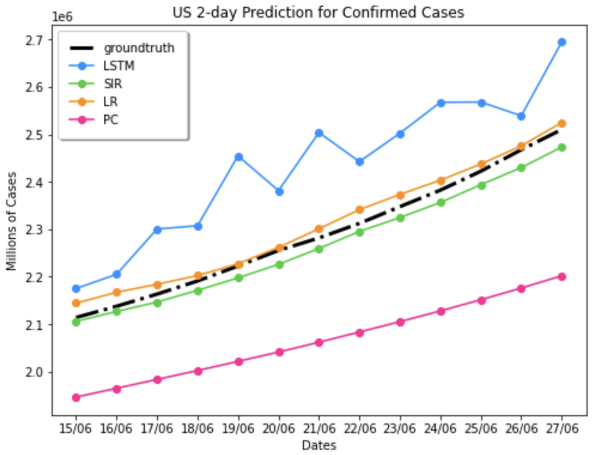
3 Node Dense (output)

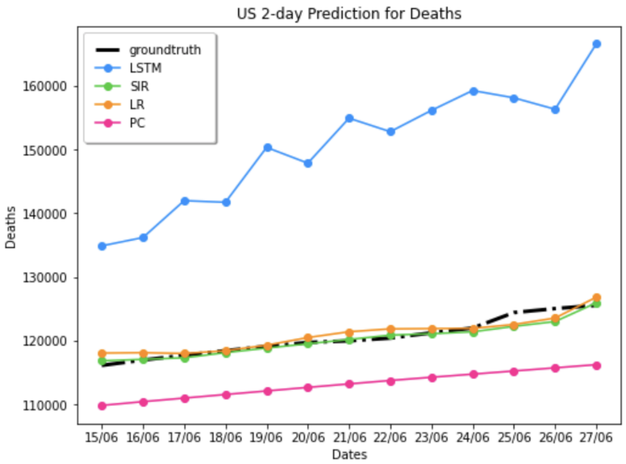


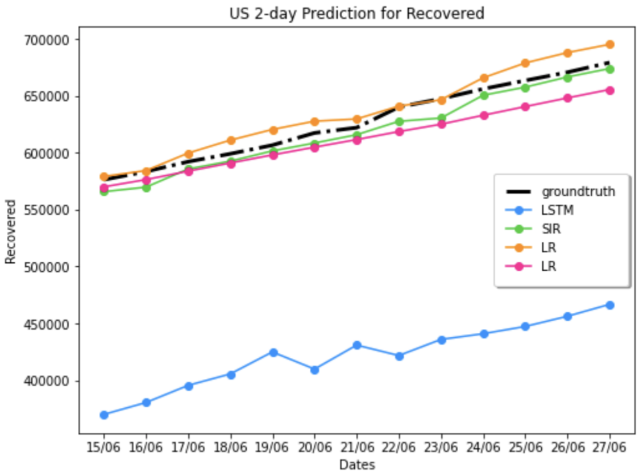




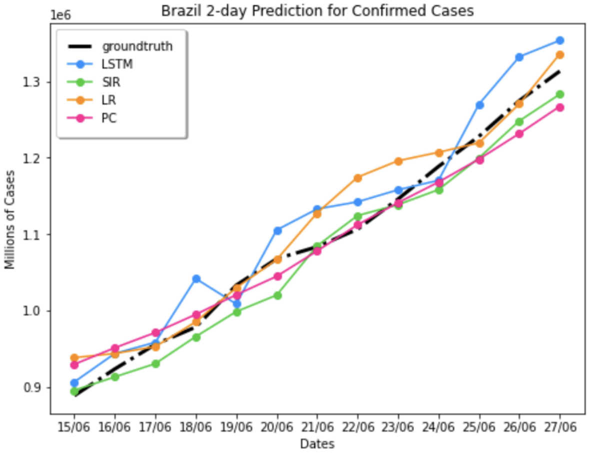
2-day Predictions: US

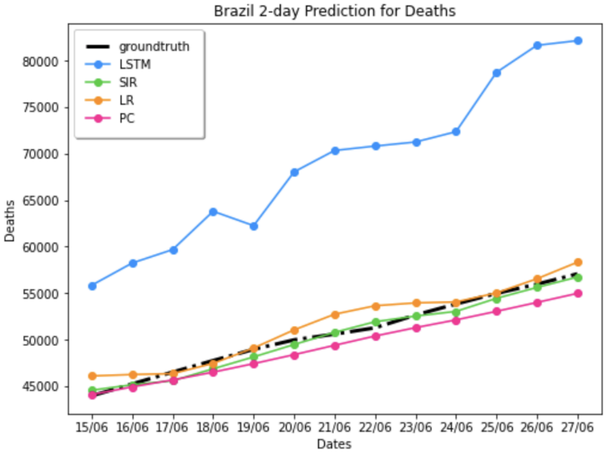


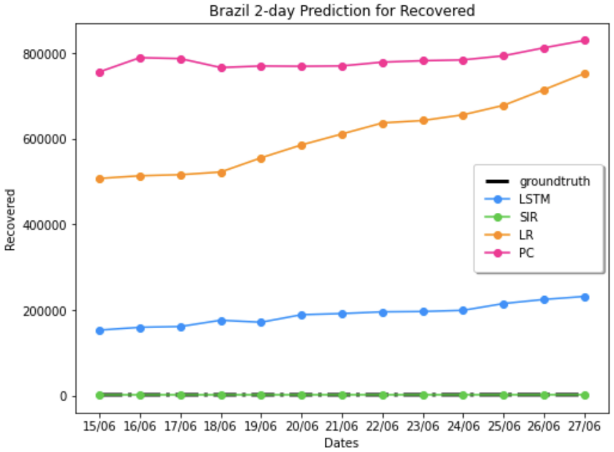




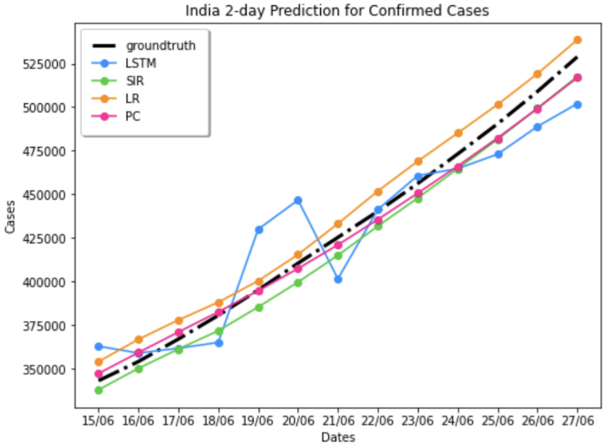
2-day Predictions: Brazil

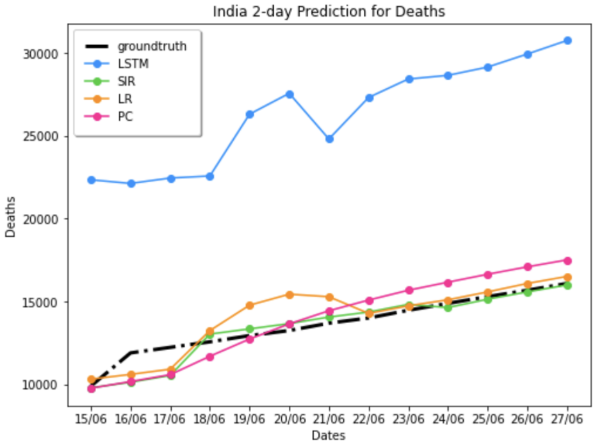


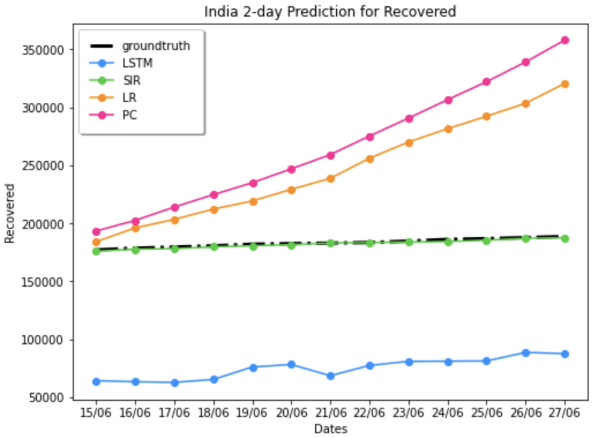




2-day Predictions: India



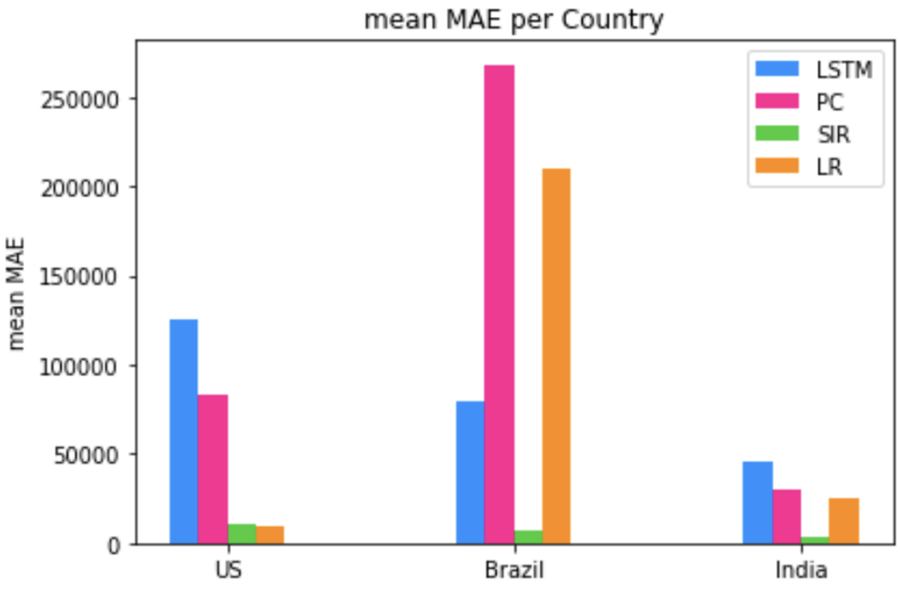


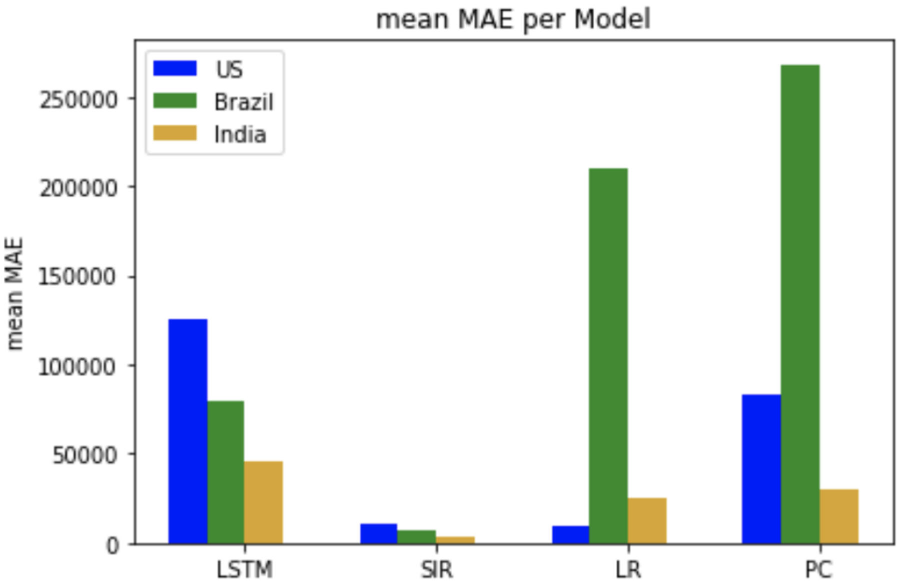


2-day Predictions: Evaluation

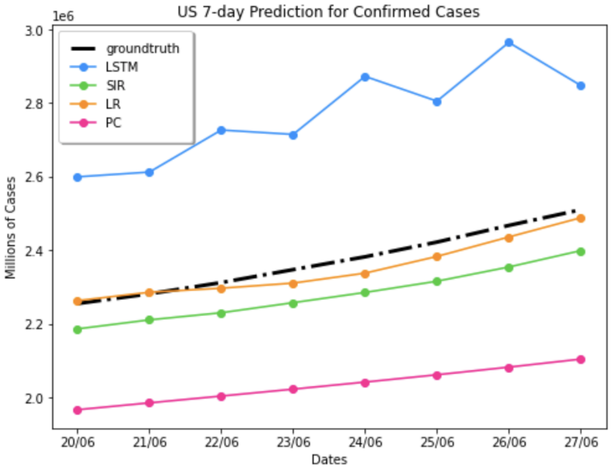
Mean Absolute Error:

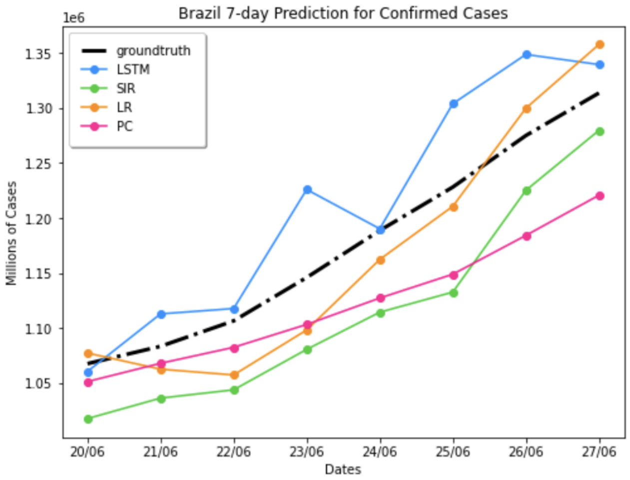
0.33 *
(abs(# predicted cases - # actual cases)
+
abs(# predicted deaths - # actual deaths)
+
abs(# predicted recovered - # actual recovered))

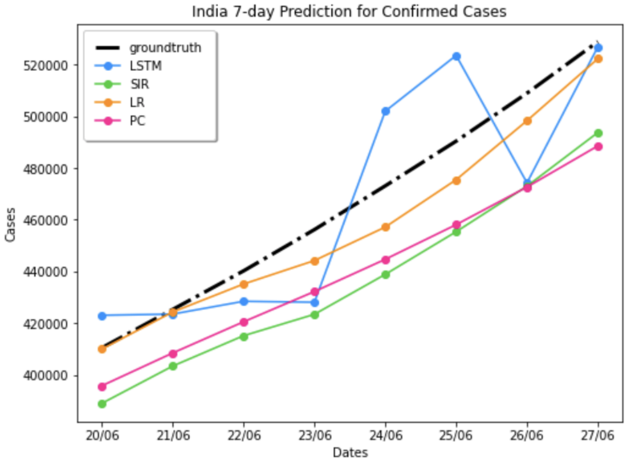




7-day Predictions



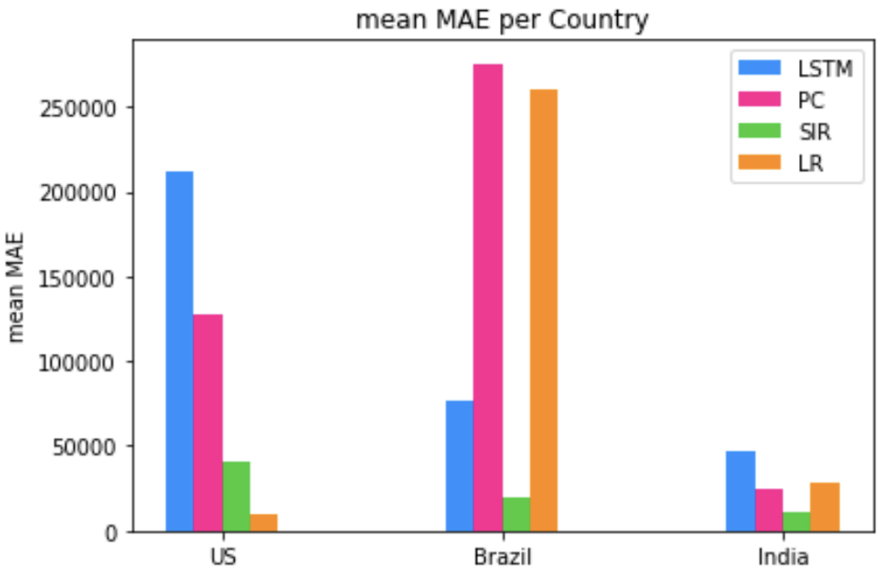


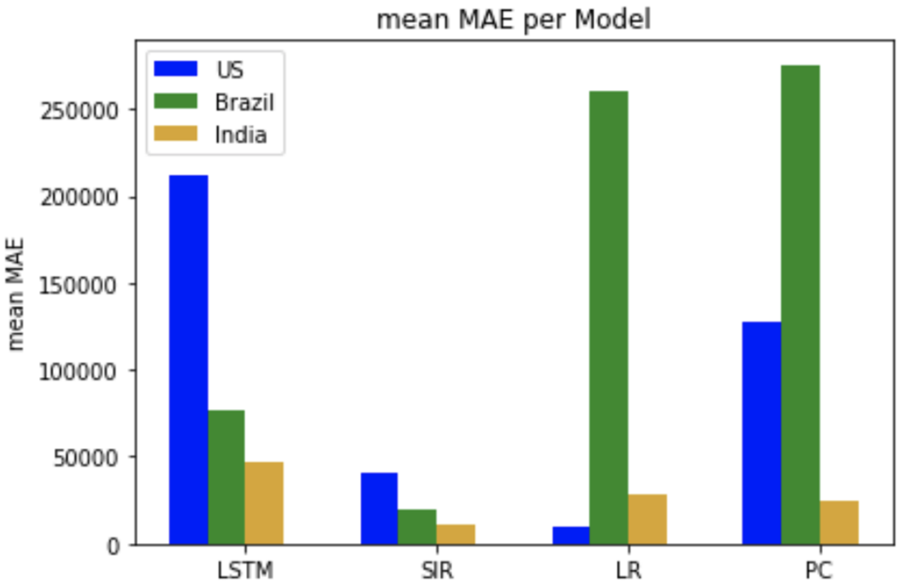


7-day Predictions: Evaluation

<u>Mean Absolute Error:</u>

0.33 *
(abs(# predicted cases - # actual cases)
+
abs(# predicted deaths - # actual deaths)
+
abs(# predicted recovered - # actual recovered))





Conclusion

Learnings:

- SIRD-model best estimate covid-19 BUT may lose interpretability
- LSTM with multiple outputs need well prepared training data
- Some 7day models may be more accurate for 2day predictions
- Complex models have a lot more effort to get robust results.

Work available at: https://gitlab.ethz.ch/olemuell/pandeml