

Krótkoterminowa predykcja rozwoju pandemii

COVID-19



AUTORZY

Wojciech Konieczkowicz
Jakub Szczudło
Karol Talaga
Michał Wełna
Dominik Woźniak

OPIEKUNOWIE

dr inż. Waldemar Bauer
dr Katarzyna Grobler-Dębska

Spis treści

1

Zestawy danych

Dane, na podstawie których dokonano predykcji

2

Wykorzystane metody

Prophet, ARIMAX, LSTM

3

Podsumowanie

Komentarz do otrzymanych wyników

Zestawy danych

Prophet

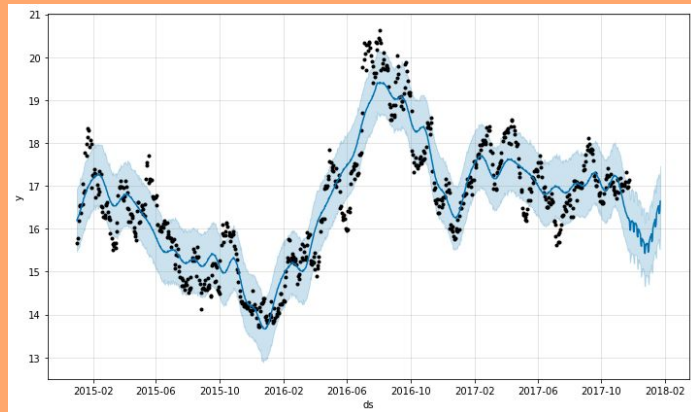
Open Disease Data
(disease.sh)

LSTM i ARIMA

wybrane cechy z arkusza danych
Michała Rogalskiego
(bit.ly/covid19-poland)

Prophet

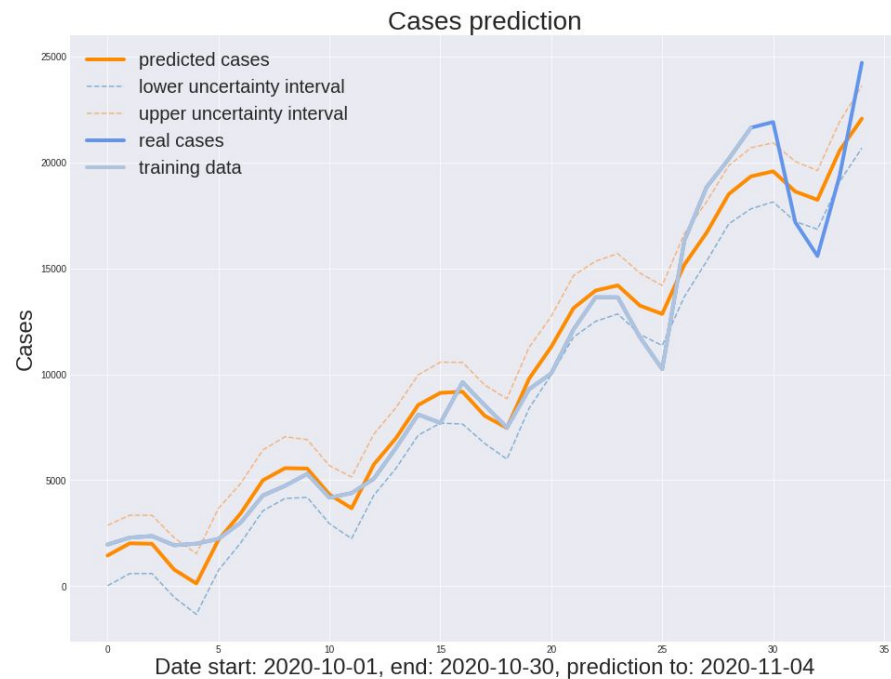
- Biblioteka stworzona przez Facebooka do prognozowania szeregów czasowych.
- Najlepiej sprawdza się w przypadkach, które mają silne efekty sezonowości.
- Odporność na brak danych oraz na występowanie wartości odstających



Cechy:

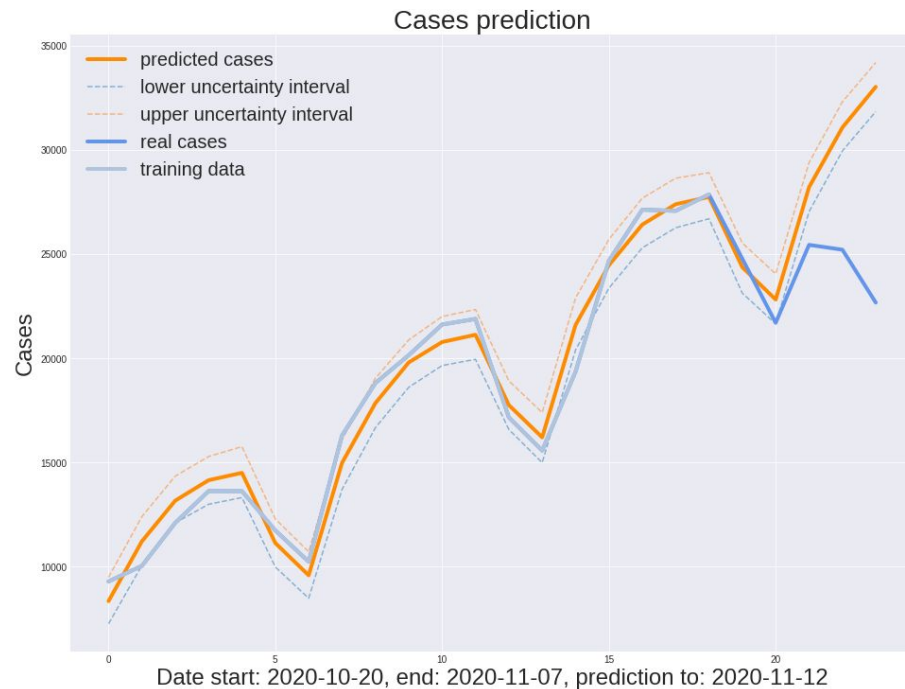
- Szybkość działania i dokładność
- Całkowicie automatyczny
- Możliwość dostrojenia

Prophet - wyniki



Doskonale widać skłonności Prophetą do prognozowania trendu.

Prophet - wyniki

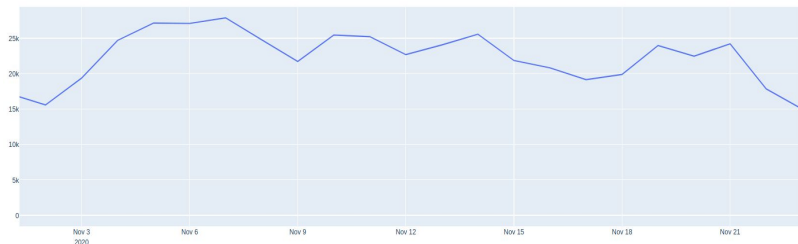


Doskonale widać skłonności Prophet'a do prognozowania trendu - z tego powodu nie poradził sobie on z przewidzeniem nagłego spadku liczby pozytywnych przypadków.

ARIMAX

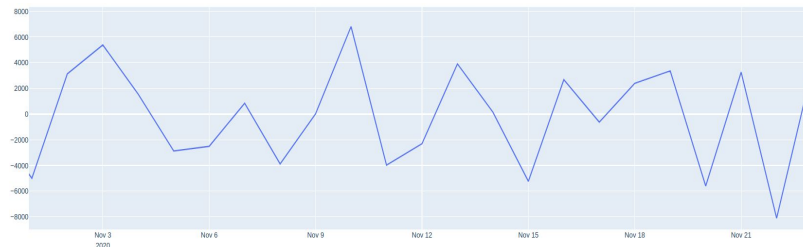
ARIMA (Autoregressive Integrated Moving Average)

$$\Delta y_t = \sum_{i=1}^p \phi_i \Delta y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$



niestacjonarny

$$\Delta y_t = y_t - y_{t-1}$$



stacjonarny

ARIMAX

ARIMA (Autoregressive Integrated Moving Average)

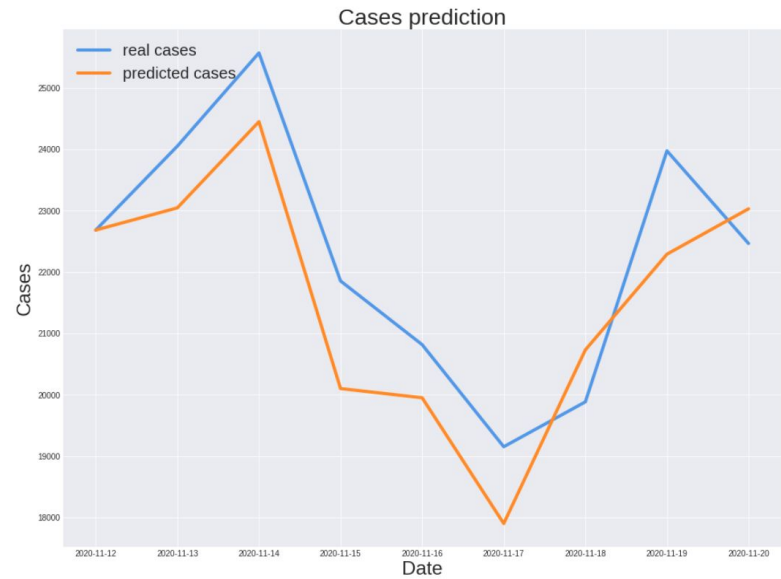
$$\Delta y_t = \sum_{i=1}^p \phi_i \Delta y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

ARIMAX (eXogenous variable)

$$\Delta y_t = \sum_{i=1}^p \phi_i \Delta y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^r \beta_k X_{t-k} + \epsilon_t$$



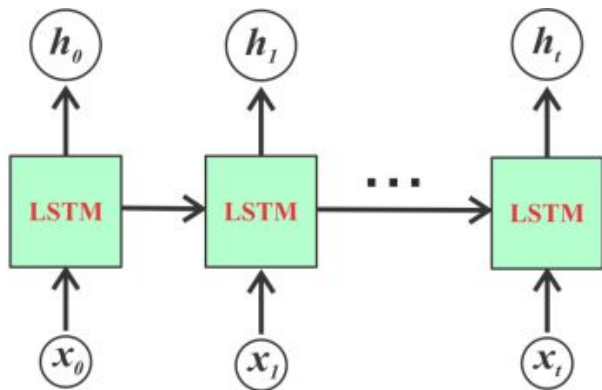
ARIMAX - wyniki



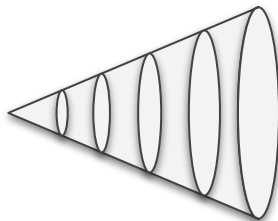
Model potrafi dobrze przewidzieć wartości przypadków nawet dla 9 dni.

LSTM

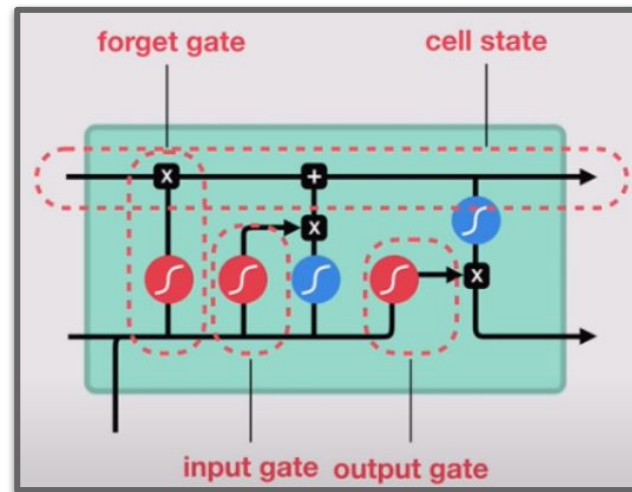
- Long Short-Term Memory
- często stosowany dla danych sekwencyjnych
- jego budowa zapobiega problemom innych tego typu modeli



codeproject.com



Illustrated Guide to LSTM's and GRU's:
A step by step explanation [youtube.com]



LSTM - przygotowania

	cases	cases_per_mln_cum	deaths	recovered	tests	positives	positives_tests_ratio	hospitalized	quaran	supervised	deaths_recovered_ratio
day											
2020-11-14	25571	18007.2	548	14635	46607	21854	0.468900	21988	409322	29920	0.037444
2020-11-15	21854	18576.7	303	12568	35105	20816	0.592964	22458	422767	28528	0.024109
2020-11-16	20816	19119.0	143	11239	41983	19152	0.456185	23033	415174	25618	0.012724
2020-11-17	19152	19618.0	357	18260	57012	19883	0.348751	22812	415601	25653	0.019551
2020-11-18	19883	20136.1	603	18601	58011	23975	0.413284	22536	434034	24717	0.032418
2020-11-19	23975	20760.8	637	19003	55048	22464	0.408080	22643	408567	23153	0.033521
2020-11-20	22464	21346.1	626	19914	43945	24213	0.550984	22478	391837	22982	0.031435
2020-11-21	24213	21976.9	574	25275	42860	18467	0.430868	22094	356130	22156	0.022710
2020-11-22	17856	22442.2	330	16896	27498	15002	0.545567	22234	362959	21290	0.019531
2020-11-23	15002	22833.1	156	14897	33717	10139	0.300709	22405	334331	19671	0.010472

surowe dane

LSTM - przygotowania

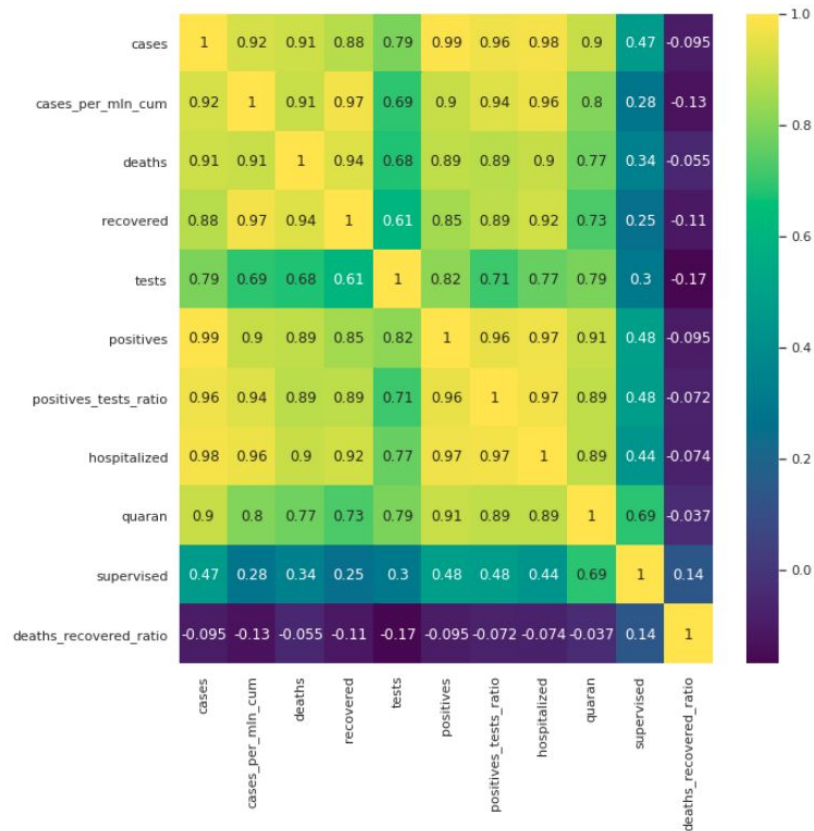
- 1) analiza korelacji między predyktorami oraz między predyktorami a zmienną docelową
- 2) zastosowaniu algorytmu *greedy elimination*
- 3) przekształcenie wartości cech dla danego dnia na różnice między następującymi sobie dniami

$$\Delta y_t = y_t - y_{t-1}$$

- 4) transformacja cech do rozkładu normalnego
- 5) standaryzacja

$$Z = \frac{X - \mu}{\sigma}$$

macierz korelacji

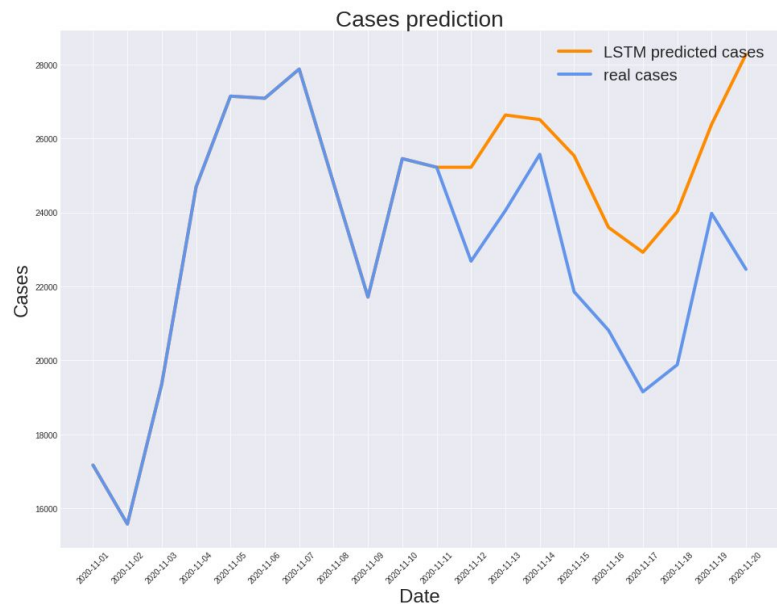


LSTM - przygotowania

	cases	deaths	tests	hospitalized_sqrt	quaran	supervised
2020-10-03	-0.156901	-0.055126	-0.474882	-0.731864	-0.467028	0.136411
2020-10-04	-0.393647	-0.213397	-0.782027	-0.550560	-0.037052	-0.176946
2020-10-05	-0.158299	-0.097331	0.536532	0.540572	0.015012	-0.395517
2020-10-06	-0.084666	0.177005	1.948230	-0.235605	-0.702900	0.355788
2020-10-07	0.165594	0.050388	-0.021418	-0.694212	0.785980	0.380491

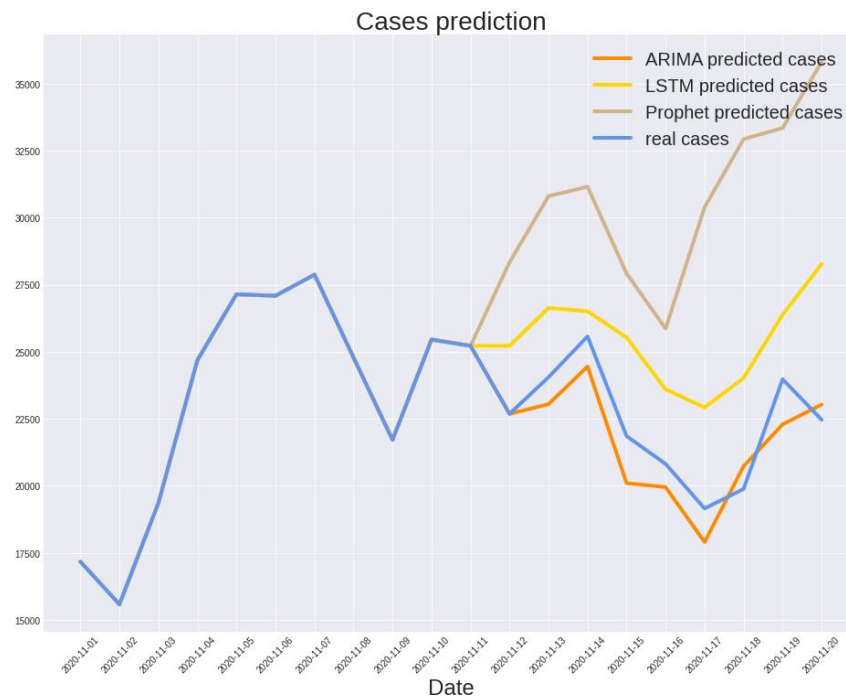
przetworzone dane

LSTM - wyniki



Przybliżona predykcja oparta na różnicach -
model o najmniejszym znalezionym błędzie walidacyjnym.

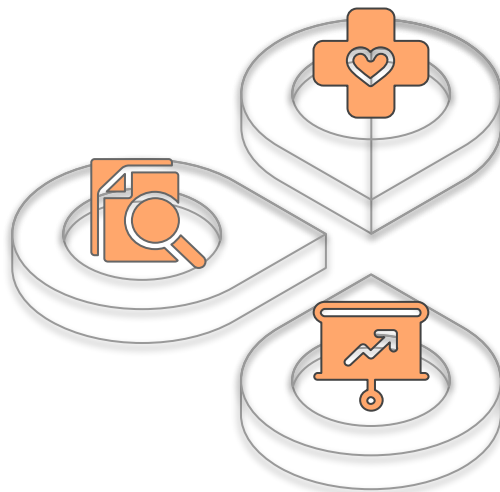
Porównanie metod



Podsumowanie

DANE I NIEPEWNOŚĆ

Modele zostały oparte o dane zależne od sprawności funkcjonowania instytucji publicznych oraz liczby wykonanych testów



PRZYDATNOŚĆ

Wyniki obrazują wstępną przydatność

CO DALEJ?

Analiza wpływu wydarzeń społecznych na rozwój pandemii, poprawa modeli ARIMAX i LSTM

Credits

Presentation template by [Slidesgo](#)

Icons by [Flaticon](#)

Images & infographics by [Freepik](#)

Author introduction slide photo created by **katemangostar** - Freepik.com

Big image slide photo created by **jcomp** - Freepik.com

Text & Image slide photo created by **rawpixel.com** - Freepik.com

Text & Image slide photo created by **Freepik**