

Światowy program szczepień przeciwko COVID-19

Eksploracja Danych

Marek Grudkowski
Kamil Kaczmarkiewicz

Politechnika Gdańska

- zbiór danych opisujący postęp światowego programu szczepień przeciwko COVID-19
- dane pochodzą z wielu źródeł (głównie organy krajowe)
- zbiór jest ciągle aktualizowany - od momentu rozpoczęcia prac liczba przykładów wzrosła z 13 na 20 tys
- jeden przykład jest opisany za pomocą 15 atrybutów, które w większości są numeryczne

- wskazać państwa, które radzą sobie najlepiej w programie szczepień, by inne kraje mogły się na ich działaniach wzorować
- oszacować zapotrzebowanie w szczepionkach na nadchodzące miesiące, by uniknąć takich sytuacji jak brakujące czy marnujące się jej dawki (dokładność ok. 80%)
- oszacować teoretyczną datę uzyskania przez dane państwo odporności zbiorowej (dokładność ok. 80%)

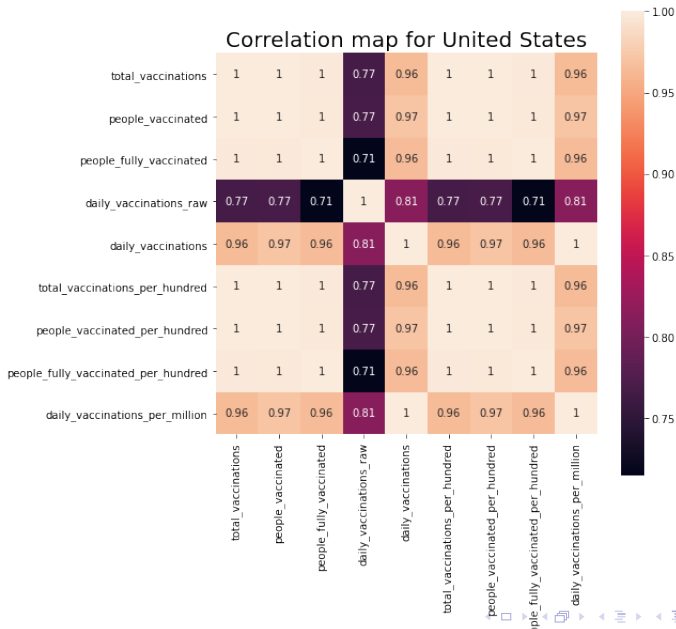
Przygotowanie danych

Size of data is: (13307, 13)

Missing values in dataset:

country	0
iso_code	0
date	0
total_vaccinations	5255
people_vaccinated	5931
people_fully_vaccinated	7926
daily_vaccinations_raw	6529
daily_vaccinations	220
total_vaccinations_per_hundred	5255
people_vaccinated_per_hundred	5931
people_fully_vaccinated_per_hundred	7926
daily_vaccinations_per_million	220
vaccines	0

Przygotowanie danych



Państwa wiodące (stan na 25.05)

2 Izrael (59%)

Państwa wiodące (stan na 25.05)

- 2 Izrael (59%)
- 4 Stany Zjednoczone (39%)

Państwa wiodące (stan na 25.05)

- 1 Zjednoczone Emiraty Arabskie (75%)
- 2 Izrael (59%)
- 3
- 4 Stany Zjednoczone (39%)

Państwa wiodące (stan na 25.05)

- 1 Zjednoczone Emiraty Arabskie (75%)
- 2 Izrael (59%)
- 3 Chile (42%)
- 4 Stany Zjednoczone (39%)

Państwa wiodące (stan na 25.05)

- 1 Zjednoczone Emiraty Arabskie (75%)
- 2 Izrael (59%)
- 3 Chile (42%)
- 4 Stany Zjednoczone (39%)
- 5 Chiny (37%)

Państwa wiodące (stan na 25.05)

- 1 Zjednoczone Emiraty Arabskie (75%)
- 2 Izrael (59%)
- 3 Chile (42%)
- 4 Stany Zjednoczone (39%)
- 5 Chiny (37%)
- 6 Polska (16%)

Regresja



Potężne
narzędzie matematyczne

imgflip.com

Sieć neuronowa



Przetrenowałam się

Sieć neuronowa



Czego mam się
dzisiaj nauczyć?

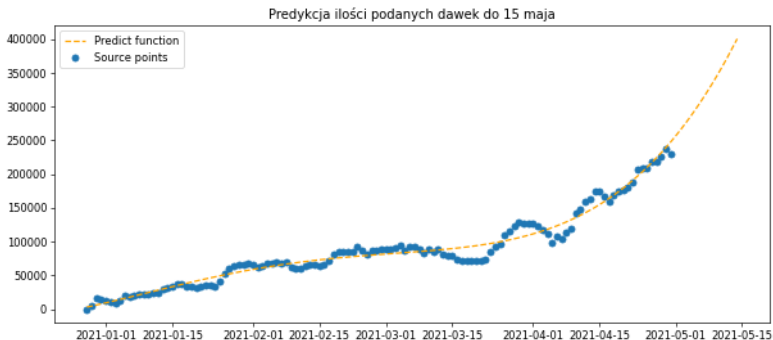
imgflip.com

Regresja liniowa



Wielomian trzeciego
stopnia? Oszalałeś?

Skuteczność regresji



Rysunek: Ambitna predykcja programu szczepień w Polsce

- dużo pracy wymagało wypełnienie brakujących wartości
- algorytm przygotowywania danych można wykorzystać w przyszłości
- łatwo można wskazać liderów w programach szczepień
- algorytm regresji nie jest tutaj dobrym rozwiązaniem
- należy zbudować lepszy, bardziej złożony model