COVID-19 AI challenges on Kaggle

Anthony Goldbloom, CEO of Kaggle
April 1, 2020

Twitter: @antgoldbloom

Kaggle is the world's largest machine learning community with 4.4MM members

Twitter: @antgoldbloom

kaggle

We are running three COVID-19 challenges

1. Automated literature review: NLP challenge on 45K academic papers
2. Forecasting challenge: forecast cases and fatalities by city
3. Dataset challenge: sharing datasets useful for making decisions on aspects of the pandemic

kaggle

We are running three COVID-19 challenges

1. **Automated literature review: NLP challenge on 45K academic papers**
2. Forecasting challenge: forecast cases and fatalities by city
3. Dataset challenge: sharing datasets useful for making decisions on aspects of the pandemi

kaggle

https://github.com/gsarti/covid-papers-browser

```
repr_strings = [r'\br0\b', r'$r_0$', r'\br 0\b',
                'reproduction number', 'reproduction rate',
                'reproductive number', 'rate of reproduction']
```

Table of Reproduction Rates (R / R_0)

publication_date	authors	title_link	key_passages
2020-01-19	Tianmu Chen et al	A mathematical model for simulating the transmission of Wuhan novel Coronavirus (nan)	*Failed to extract figures - check manually.*
2020-01-24	Shi Zhao et al	Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak (nan)	In Table1, we estimated that the R0 ranges from 2.24 (95%CI: 1.96-2.55) to 5.71 (95%CI: 4.24-7.54) associated with 8-fold to 0-fold increase in the reporting rate We estimated the mean R0 of 2019-nCoV ranging from 2.24 (95%CI: 1.96-2.55) to 3.58 (95%CI: 2.89-4.39) if the reporting effort has been increased by a factor of between 8-and 2-fold after the diagnostic protocol released on January 17, 2020 and many medical supplies reached Wuhan. https://doi.org/10.1101/2020.01.23.916395 doi: bioRxiv preprint Table Table 1 The summary table of the estimated basic reproduction number, R0, under different scenarios
		Pattern of early human-to-human	

Source: https://www.kaggle.com/davidmezzetti/cord-19-transmission-incubation-environment

Transmission, incubation, and environmental stability

Range of incubation periods for the disease in humans (and how this varies across age and health status) and how long individuals are contagious, even after recovery.

Range of incubation periods for the disease in humans

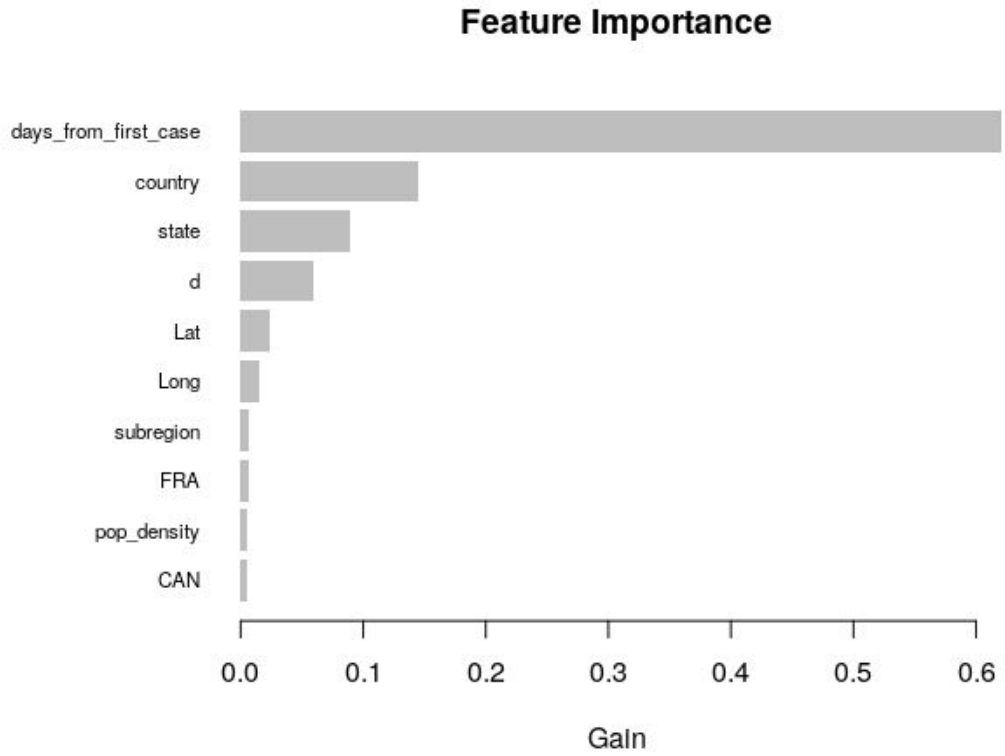
Date	Study	Days	Range (Days)
2020-01-28	Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data	5.6	
2020-02-04	The incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and application	5.2	
2020-02-13	Serial interval of novel coronavirus (2019-nCoV) infections	5	
2020-02-20	SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration	3	
2020-02-23	Epidemiological characteristics of 1212 COVID-19 patients in	4.75	50% between 3

Source: https://www.kaggle.com/covid-19-contributions

We are running three COVID-19 challenges

1. Automated literature review: NLP challenge on 45K academic papers
2. **Forecasting challenge: forecast cases and fatalities by city**
3. Dataset challenge: sharing datasets useful for making decisions on aspects of the pandemic

kaggle



Source: https://www.kaggle.com/nonserial/covid19-prediction-with-restcountries-eu-api-data

kaggle

We are running three COVID-19 challenges

1. Automated literature review: NLP challenge on 45K academic papers
2. Forecasting challenge: forecast cases and fatalities by city
3. **Dataset challenge: sharing datasets useful for making decisions on aspects of the pandemic**

kaggle

High Temperature and High Humidity Reduce the Transmission of COVID-19

Jingyuan Wang, Ke Tang, Kai Feng and Weifeng Lv*

March 9, 2020

Abstract. This paper investigates how air temperature and humidity influence the transmission of COVID-19. After estimating the serial interval of COVID-19 from 105 pairs of the virus carrier and the infected, we calculate the daily effective reproductive number, R, for each of all 100 Chinese cities with more than 40 cases. Using the daily R values from January 21 to 23, 2020 as proxies of non-intervened transmission intensity, we find, under a linear regression framework for 100 Chinese cities, high

Source: https://www.kaggle.com/nonserial/covid19-prediction-with-restcountries-eu-api-data

kaggle

Country/Region	Lat	Long	Date	Confirmed Cases	Fatalities	day_from_jan_first	temp	min	max
Afghanistan	33.0	65.0	2020-01-22	0.0	0.0	22	42.6	33.6	54.9
Afghanistan	33.0	65.0	2020-01-23	0.0	0.0	23	42.0	32.7	55.9
Afghanistan	33.0	65.0	2020-01-24	0.0	0.0	24	40.1	36.9	43.2
Afghanistan	33.0	65.0	2020-01-25	0.0	0.0	25	46.0	37.9	56.3
Afghanistan	33.0	65.0	2020-01-26	0.0	0.0	26	42.8	36.1	53.1
Afghanistan	33.0	65.0	2020-01-27	0.0	0.0	27	43.0	36.5	50.7
Afghanistan	33.0	65.0	2020-01-28	0.0	0.0	28	41.7	34.7	48.2
Afghanistan	33.0	65.0	2020-01-29	0.0	0.0	29	15.2	13.3	16.9
Afghanistan	33.0	65.0	2020-01-30	0.0	0.0	30	15.2	13.3	16.9

Source: https://www.kaggle.com/davidbnn92/weather-data/output

kaggle

Challenges: www.kaggle.com/covid19

Takeaways: www.kaggle.com/covid-19-contributions

Wanted: epidemiologists, virologists, public health volunteers

a@kaggle.com

Twitter: @antgoldbloom

kaggle