Network Modeling Analysis in Health Informatics and Bioinformatics (2022) 11:14
https://doi.org/10.1007/s13721-022-00359-1

SHORT COMMUNICATION

Deathdaily: A Python Package Index for predicting the number of daily COVID-19 deaths

Yoshiyasu Takefuji[1]

Received: 9 December 2021 / Revised: 7 February 2022 / Accepted: 5 March 2022 / Published online: 20 March 2022

Abstract

This paper presents a new open-source program called deathdaily for predicting the number of daily COVID-19 deaths in the next 7 days. The predictions can be used by policymakers to determine whether current policies should be strengthened/mitigated or new policies should be challenged to mitigate the COVID-19 pandemic. Although vaccines have been mitigating the pandemic initially, the recent resurgence with new variants has been observed in many vaccinated countries. This paper shows how to use deathdaily to detect symptoms of resurgence. The proposed deathdaily is available in public and can be installed by a Python package manager PyPI. The deathdaily has been downloaded by 15,964 users worldwide, according to https://pepy.tech/project/deathdaily. The fact shows that the applicability, practicality, and usefulness of the proposed program have been duly evaluated.

Keywords COVID-19 · The number of daily deaths · Scraping a dataset · Vaccines · Variants

1 Introduction

The number of daily infections has been used by policymakers to determine whether current policies should be strengthened or mitigated. The ultimate goal of COVID-19 policy is to reduce the number of daily deaths due to COVID-19 without weakening economic activity. Therefore, the policy performance should be daily evaluated for updating the current policy. However, the traditional tools that predict the number of people infected (cases) each day do not provide a good indicator for policymakers because the number will not be correct unless every person is tested multiple times each day.

R, the reproduction number with the number of cases has been used by many policymakers worldwide (Kucharski 2020; Zhang et al. 2020; Caicedo-Ochoa 2022). R is the average number of people who contract a disease and are subsequently infected (David Adam 2020). Therefore, R is an imprecise estimate that rests on assumptions (David Adam 2020; Maruotti et al. 2021; Shaw and Kennedy 2021).

This paper proposes a good indicator of the pandemic by the number of daily deaths instead of the number of cases. The number of daily deaths due to COVID-19 is a really good indicator for policymakers (Takefuji 2021). However, the average lag between daily COVID-19 cases and deaths was 8 days (Jin 2021).

This paper presents a new open-source tool, dailydeath that can serve as a good indicator for policymakers by predicting the number of daily deaths for the next seven days. When using the proposed tool, policymakers need to consider an eight-day lag with a standard deviation of 4 days (Jin 2021).

The Python program, deathdaily can generate a graph of a country on the number of daily deaths due to COVID-19 where it is composed of two lines: a black colored line as true number and a blue colored line for prediction with nth-degree polynomial regression. The legend in the generated graph contains important parameters: the country name, the number of days to be used for prediction, the nth-degree polynomial regression, and R-squared (r_2). r_2 is commonly used as a statistical measure of how close the data is to the fitted regression line.

The prediction from the latest day of data to 7 days ahead using historical data is based on a curve fitting function. The curve fitting function requires two determinants, the size

✉ Yoshiyasu Takefuji
takefuji@keio.jp

[1] Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

of the data (number of days) and the "*n*" of the *n*th degree polynomial regression.

The latest data on daily death is automatically scraped and downloaded from the following jhu web site: https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/jhu/new_deaths.csv

The user needs to provide three determinants to the deathdaily tool: the name of the country, the size of the data (days), and the "*n*" of the *n*th degree polynomial. The higher the degree polynomial, the better the curve fitting will be. However, curve overfitting does not necessarily mean good prediction. The proposed deathdaily can show the trend of daily deaths in the very near future by the user changing the size of days and the *n*th degree polynomial. The user is allowed to change three determinants to observe the calculated prediction.

R-squared (r_2) can let you know if n in an nth order polynomial should be smaller or not. The data size can also affect the curve fitting of the polynomial regression. Due to the nature of curve fitting, data from the most recent days may have a stronger impact on prediction than data from older days. The larger the data size and the smaller the polynomial degree, the more underfitting the prediction. Conversely, the smaller the data size and the larger the polynomial degree, the more the prediction is overfitting. This is the result of prediction using two determinants, data size and polynomial degree, and validation is up to the user. There is no automated algorithm for finding the optimal determinants. If an automated algorithm existed, future states could be predicted.

Deathdaily is available in public and can be easily installed by the PyPI package. According to PyPI stats with https://pepy.tech/project/deathdaily, the deathdaily has been downloaded by 15,964 users worldwide. This fact shows that its applicability and usefulness was justified.

2 Deathdaily

According to PYPL (2022) and TIOBE (2022), Python is the best open-source development language. The proposed deathdaily is written in Python. The deathdaily.py program consists of two modules: one that scrapes the Internet for the most recent data set on the number of daily deaths, and another with an *n*th-degree polynomial regression given to predict the number of daily deaths for the next 7 days.

In statistics, polynomial regression is a form of regression analysis that models the relationship between an independent variable x and a dependent variable y as an n^{th} degree polynomial in x. Polynomial regression is commonly used to observe trends and tendencies in a variety of applications(Zhang and Jiang 2021; Lee et al. 2011; Liu et al. 2019; Davies et al. 2021). In deathdaily.py as shown in Fig. 1, polynomial regression is implemented by polyfit function and poly1d function in numpy library (np):

model = np.poly1d(np.polyfit(*x*[valid],*y*[valid],degree)).

np.poly1d is a one-dimensional polynomial class. The *n*th-degree polynomial regression model can be built. The prediction y can be given by *y* = model(*x*) where *x* is an independent variable as shown in Fig. 1.

For r-squared calculation, sklearn.metrics library is used where *r*2_score function can give the value of r-squared as shown in Fig. 1.

from sklearn.metrics import *r*2_score

In the data preprocessing, the new_deaths.csv file contains an erroneous minus sign (-), so we remove the minus signs with the following Python commands:

sp.call("cat new_deaths.csv | sed '2,$s/-//g' > new",shell = True).

sp.call("mv new new_deaths.csv",shell = True).

data = pd.read_csv("new_deaths.csv").

3 Results of running deathdaily

Dataset new_deaths.csv on daily deaths due to COVID-19 can be scraped from the following site: https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/jhu/new_deaths.csv

In order to run deathdaily, you must install it by the following PyPI packaging command:

$ pip install deathdaily

The author picked countries such as Germany, the US, France, and the UK to justify the proposed claim.

Then, run deathdaily

$ deathdaily Germany 200 5

The above command is to produce a graph of Germany using 200 days from the executed day with the 5th-degree polynomial regression.

Figure 2 shows the result on Germany. Germany is observed to be in the midst of resurgence. Germany is 68.9% fully vaccinated as of Dec.6, 2021.

Figure 3 shows the strong resurgence in the US. Figure 3 was generated by the following command:

$ deathdaily 'United States' 2005

The US is 59.9% fully vaccinated as of Dec.6, 2021. The pandemic situation in the US is getting worse.

Figure 4 shows that France is observed to be in the midst of resurgence with 70.2% fully vaccinated.

Figure 5 depict that the UK is observed to be in the midst of resurgence with 69.2% fully vaccinated.

```
import pandas as pd
import numpy as np
import sys
from time import sleep
import matplotlib.pyplot as plt
import subprocess as sp
from sklearn.metrics import r2_score as r2
import matplotlib.patches as mpatches

sp.call("wget https://raw.githubusercontent.com/owid/covid-19-
data/master/public/data/jhu/new_deaths.csv",shell=True)
sp.call("cat new_deaths.csv|sed '2,$s/,-/,/g' >new",shell=True)
sp.call("mv new new_deaths.csv",shell=True)
data=pd.read_csv("new_deaths.csv")
sp.call("rm new_deaths.csv",shell=True)

class main:
 def main(self,country,days=400,degree=7):
  n=len(data[country])
  y=data[country][n-days:n]
  for i in y:
   print(i)
  x=np.arange(n-days,n)
  valid = ~(np.isnan(x) | np.isnan(y))
  model=np.poly1d(np.polyfit(x[valid],y[valid],degree))
  date=data['date'][n-1]
  x1=np.arange(n-days,n+7)
  y1=model(x1)
  ny1=[]
  for i in y1:
   if i<0:i=0
   ny1.append(i)
  x2=np.arange(n-days,n)
  y2=model(x2)
  r2s=round(r2(y,y2),3)
  plt.plot(x,y,'k')
  plt.plot(x1,ny1,'b')
  ax=plt.subplot()
  handles,labels = ax.get_legend_handles_labels()
  st='daily deaths in '+str(country)+'\n'+str(days)+' days from
'+str(date)+'\n'+str(degree)+'th regression\n'+'r2: '+str(r2s)
  handles.append(mpatches.Patch(color='none', label=st))
  plt.legend(handles=handles)
  plt.savefig(country+".png")
  plt.show()
country=""
days=400
degree=5
if len(sys.argv)==1:
 print('country name is needed!')
 sys.exit()
if len(sys.argv)==2:
 if sys.argv[1] in data.columns:
  country=str(sys.argv[1])
 else:
  print('correct country name!')
  sys.exit()
if len(sys.argv)==3:
 if sys.argv[1] in data.columns:
  country=str(sys.argv[1])
  if int(sys.argv[2])>len(data[country]):
   print('use smaller days')
   sys.exit()
  else:
   days=int(sys.argv[2])
 else:
  print('correct country name')
  sys.exit()
if len(sys.argv)==4:
 if sys.argv[1] in data.columns:
  country=str(sys.argv[1])
  if int(sys.argv[2])>len(data[country]):
   print('use smaller days')
   sys.exit()
  else:
   days=int(sys.argv[2])
   if int(sys.argv[3]) > 4:
    degree=int(sys.argv[3])
   else:
    print('use higher degree number')
    sys.exit()
 else:
  print('correct country name')
  sys.exit()
m=main()
m.main(country=country,days=days,degree=degree)
```

Fig. 1 Deathdaily.py Python program

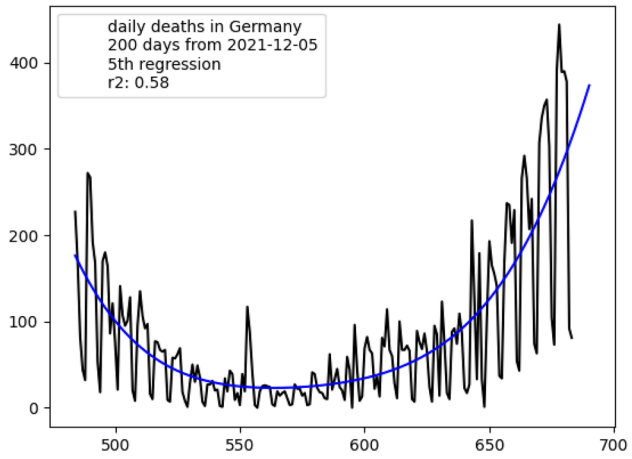

Fig. 2 Result with deathdaily Germany 2005

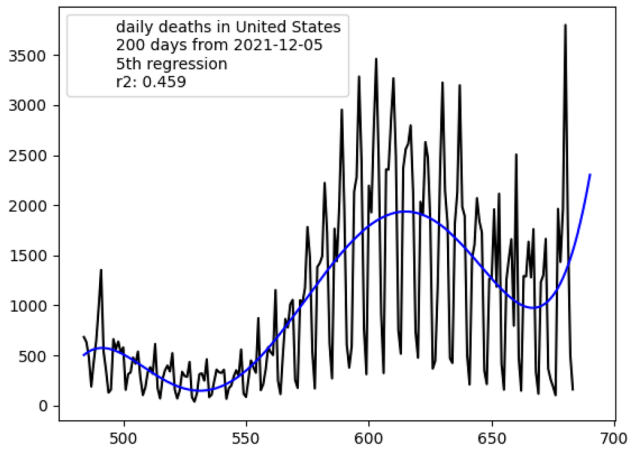

Fig. 3 Result with deathdaily 'United States' 2005

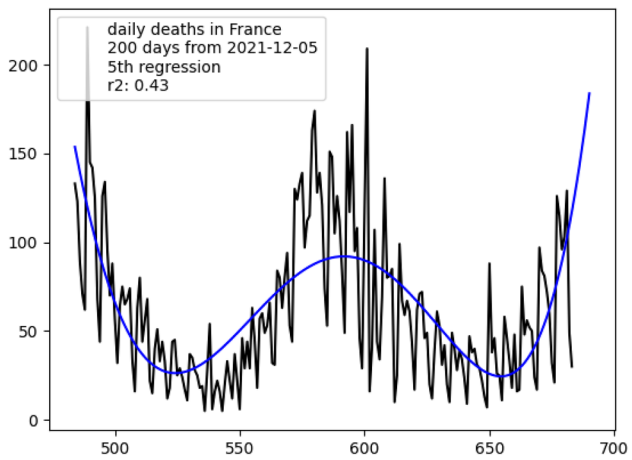

Fig. 4 Result with deathdaily France 2005

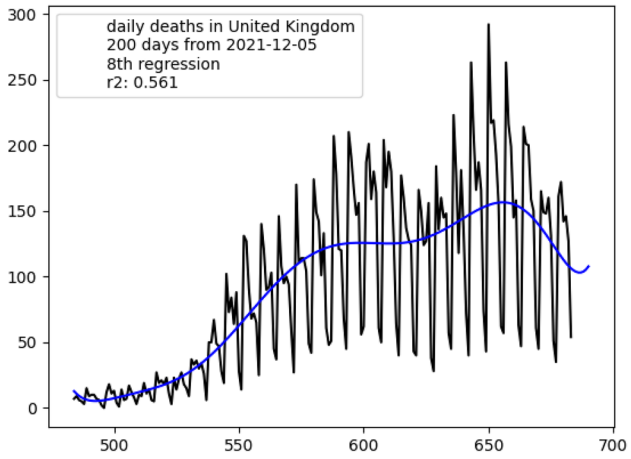

Fig. 5 Result with deathdaily 'United Kingdom' 2008

4 Discussion

The number of daily deaths due to COVID-19 clearly shows the pandemic situation. We need to use deathdaily to detect symptoms of resurgence. The proposed program is not only easy to install by PyPI packaging, but also provides the number of daily deaths in a country and its forecast for the week ahead. Although investigated countries except the US with 59.9% are near 70% fully vaccinated as of Dec.6, 2021, symptoms of resurgence are observed in vaccinated countries. Policymakers should be able to determine whether the current policy should be strengthened or updated. Or they should challenge a new policy with excellent countries suppressing the pandemic such as New Zealand, Taiwan, Australia, South Korea, and Iceland (Takefuji 2021a). These countries use digital fences for isolating infected carriers from uninfected individuals. Scoring individual policies is reported for poor policymakers to learn the good strategies from the excellent countries (Takefuji 2021b).

5 Conclusion

The proposed deathdaily is currently used by 15,964 users in the world and the first open-source software for predicting the number of daily deaths in a country. Using the deathdaily, symptoms of resurgence were observed in vaccinated countries. Policymakers should be able to determine whether the current policy should be strengthened or mitigated. Or they should challenge a new policy with excellent countries suppressing the pandemic such as New Zealand and Taiwan.

The higher the number of cases, the higher the number of deaths. In the future, the number of daily cases can be used to improve the prediction quality of deathdaily.

Funding No fund.

Declarations

Conflict of interest The author has no conflict of interest.

Research involving human participants and/or animals Not applicable.

Informed consent Not applicable.

References

Caicedo-Ochoa Y, Rebellón-Sánchez DE, Peñaloza-Rallón M et al (2020) Effective Reproductive Number estimation for initial stage of COVID-19 pandemic in Latin American Countries. Int J Infect Dis 95:316–318. https://doi.org/10.1016/j.ijid.2020.04.069

David Adam (2020) A guide to R—the pandemic's misunderstood metric. Nature 583:346–348. https://doi.org/10.1038/d41586-020-02009-w

Davies B, Lalot F, Peitz L et al (2021) Changes in political trust in Britain during the COVID-19 pandemic in 2020: integrated public opinion evidence and implications. Humanit Soc Sci Commun 8:166. https://doi.org/10.1057/s41599-021-00850-6

Jin R (2021) The lag between daily reported Covid-19 cases and deaths and its relationship to age. J Public Health Res. https://doi.org/10.4081/jphr.2021.2049

Kucharski AJ, Russell TW, Diamond C et al (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 20:553–558. https://doi.org/10.1016/S1473-3099(20)30144-4

Lee B, LeDuc P, Schwartz R (2011) Unified regression model of binding equilibria in crowded environments. Sci Rep 1:97. https://doi.org/10.1038/srep00097

Liu Y, Zhang R, Ye H et al (2019) The development of a 3D colour reproduction system of digital impressions with an intraoral scanner and a 3D printer: a preliminary study. Sci Rep 9:20052. https://doi.org/10.1038/s41598-019-56624-3

Maruotti A, Ciccozzi M, Divino F (2021) On the misuse of the reproduction number in the COVID-19 surveillance system in Italy. J Med Virol 93(5):2569–2570. https://doi.org/10.1002/jmv.26881 (**Epub 2021 Feb 19**)

PYPL (2022) PopularitY of programming language. https://pypl.github.io/PYPL.html

Shaw CL, Kennedy DA (2021) What the reproductive number R_0 can and cannot tell us about COVID-19 dynamics. Theor Popul Biol 137:2–9. https://doi.org/10.1016/j.tpb.2020.12.003 (**Epub 2021 Jan 5**)

Takefuji Y (2021) Open schools, Covid-19, and child and teacher morbidity in Sweden. N Engl J Med 384:e66. https://doi.org/10.1056/NEJMc2101280

Takefuji Y (2021a) Analysis of digital fences against COVID-19. Health Technol. https://doi.org/10.1007/s12553-021-00597-9

Takefuji Y (2021b) SCORECOVID: a Python Package Index for scoring the individual policies against COVID-19. Health Anal 1:100005

TIOBE (2022) tiobe-index. https://www.tiobe.com/tiobe-index/

Zhang J, Jiang Z (2021) A new grey quadratic polynomial model and its application in the COVID-19 in China. Sci Rep 11:12588. https://doi.org/10.1038/s41598-021-91970-1

Zhang J, Litvinova M, Liang Y et al (2020) Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science 368(6498):1481–1486. https://doi.org/10.1126/science.abb8001