Soft Computing (2022) 26:10075–10083
https://doi.org/10.1007/s00500-022-07385-1

DATA ANALYTICS AND MACHINE LEARNING

Forecasting the spread of SARS-CoV-2 in the campania region using genetic programming

Gianni D'Angelo[1] · Salvatore Rampone[2]

Accepted: 11 July 2022 / Published online: 7 August 2022

Abstract
Coronavirus disease 19 (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus, which is responsible for the ongoing global pandemic. Stringent measures have been adopted to face the pandemic, such as complete lockdown, shutting down businesses and trade, as well as travel restrictions. Nevertheless, such solutions have had a tremendous economic impact. Although the use of recent vaccines seems to reduce the scale of the problem, the pandemic does not appear to finish soon. Therefore, having a forecasting model about the COVID-19 spread is of paramount importance to plan interventions and, then, to limit the economic and social damage. In this paper, we use Genetic Programming to evidence dependences of the SARS-CoV-2 spread from past data in a given Country. Namely, we analyze real data of the Campania Region, in Italy. The resulting models prove their effectiveness in forecasting the number of new positives 10/15 days before, with quite a high accuracy. The developed models have been integrated into the context of SVIMAC-19, an analytical-forecasting system for the containment, contrast, and monitoring of Covid-19 within the Campania Region.

Keywords COVID-19 · SARS-CoV-2 · Disease spread modeling · Spread forecasting · Genetic programming · SVIMAC-19

1 Introduction

On December 31, 2019, China reported a cluster of pneumonia cases of unknown etiology in Wuhan city. On January 30, 2020, the World Health Organization (WHO) declared the new coronavirus Sars-CoV-2 outbreak in China to be a public health emergency of international concern (Gorbalenya et al. 2020).

On January 31, 2020, the Italian government proclaimed a state of emergency and implemented the first measures to contain the infection on the entire national territory (Camporesi et al. 2022).

Since then, Coronavirus disease 2019 (COVID-19) has become an unprecedented public health crisis with a major impact on the healthcare system. This impact was evident in Europe, especially in Italy (Paterlini 2020).

In particular, the Campania Region, in Southern Italy, from the data available at the beginning of 2020, has about 5,870,000 inhabitants, making it the third most populated Region in Italy and the most populated in the South. The population density is equal to 429.4 people per Km^2, the highest value at the national level. Furthermore 63.1% of the population resides in 65 centers with more than 20,000 inhabitants. This makes the Campania Region at high risk of spreading the disease and saturating the local health system. Figure 1 shows a map of the population density of the region (Tuttitalia 2020; Siniscalchi 2018).

In response, since the beginning of the pandemic, the Campania Region has adopted a preventive management approach, supporting the use of both the tools available in the study of infectious epidemiology and the new multi-disciplinary approaches based on prediction algorithms through machine learning (Kour and Gondhi 2020).

✉ Salvatore Rampone
rampone@unisannio.it

Gianni D'Angelo
giadangelo@unisa.it

[1] DI - University of Salerno, Fisciano, SA, Italy

[2] DEMM - University of Sannio, Benevento, Italy

In this paper, we describe some models of the SARS-CoV-2 spread in the territory, and a forecasting formula then integrated in the SVIMAC-19, an analytical-forecasting system for the containment, contrast, and monitoring of COVID-19 within the Campania Region (Regione Campania 2020). Namely, our goal was to predict the number of the new daily infected people at least 10/15 days in advance.

Forecasting of a pandemic can be done based on various parameters such as the impact of environmental factors, the incubation period, the impact of quarantine, age, gender, and many others (Shinde et al. 2020; Pak et al. 2020). However, not all these data are publicly available. In this study, we used only publicly available data from both Italian National Health Organization databases and Regional repositories.

1.1 Methodology

To date, many studies have tried to identify formulas and rules able to define a mathematical model of the COVID-19 spread. Although in some cases accuracy was found to be elevated, the state-of-the-art solutions make use of many data, such as governments interventions, new drugs, and so forth, and such information could be not available or reliable. As a consequence, the resultant forecasting models are often difficult to adapt to a specific area (Tu et al. 2020).

On the contrary, our approach intended to build a forecasting model by mining useful insight from the data observed over time, without taking into account any type of external information or human intervention, in the framework of inductive inference (Angluin and Smith 1983; Rampone and Russo 2012). Such technique assesses the situations of the past thereby enabling better predictions about the situation to occur in the future.

Namely, the approach used in this study relies on the so-called Evolutionary Algorithms, and in particular on the Genetic Programming (GP) (Koza 1994; Schmidt and Lipson 2009), by improving a random population of solutions (formulae) in an evolutionary way. The performance of other algorithms widely used was also valued and compared (Fix and Hodges 1951; Altman 1992; Zhang et al. 2017).

1.2 Related works

Given its massive impacts on lives globally, the COVID-19 pandemic is a major focus of research interest at present (Doornik et al. 2022) and the list of related works is necessarily incomplete.

On March 16, 2020, the White House, collaborating with research institutes and tech companies, issued a call to action for global artificial intelligence (AI) researchers for developing novel text and data-mining techniques to assist COVID-19-related research (Alimadadi et al. 2020). Several studies investigated the kinetics of coronavirus spread through human populations (Remuzzi and Remuzzi 2020; Li et al. 2020), and the basic reproductive ratio of the virus has been estimated (Anđelić et al., 2021).

Koza (1994) laid the foundations of Genetic Programming (GP) (Affenzeller et al. 2009) and since then several variations have been made (Katoch et al. 2020; D'Angelo and Palmieri 2021).

There are numerous applications of GP in the predictive field (Rampone et al. 2021; Rampone and Valente 2021). The GP application on publicly available COVID-19 data to obtain the estimation of confirmed, deceased, and recovered cases and the epidemiology curve for countries such as China, Italy, Spain, and the USA and as well as on the global scale was afforded among others by Anđelić et al. (2021) and Salgotra et al. (2020).

Del Giudice et al. (2020) implemented a regressive model investigating some consequences of the COVID-19 pandemic in the Campania Region, taking into account how the event might affect the regional activity.

1.3 Paper outline

This paper is organized as follows: In Sect. 2, we resume the method set up, the formulae obtained, the test results, and the comparisons with some alternative methods; in Sect. 3, we show the model tuning and the experimental results during the pandemic; Sect. 4 is devoted to the Discussion and Conclusions.

2 Model set up

We aimed to find a model, expressed as a set of explicit formulae, describing the number of new infected people in Campania Region (Italy) at least 10/15 days before the occurrence. More specifically, the model we intended to build should be able to perform the prediction by starting only from information on the current infected people.

3 Reference data

The initial data were taken from an officially published set of the Campania Region.[1] The data were in according to the daily national summary of health monitoring prepared

[1] https://dati.regione.campania.it/catalogo/datasetdetail/covid-19-monitoraggio-situazione-dati-di-dettaglio-relativi-alla-regione-campania

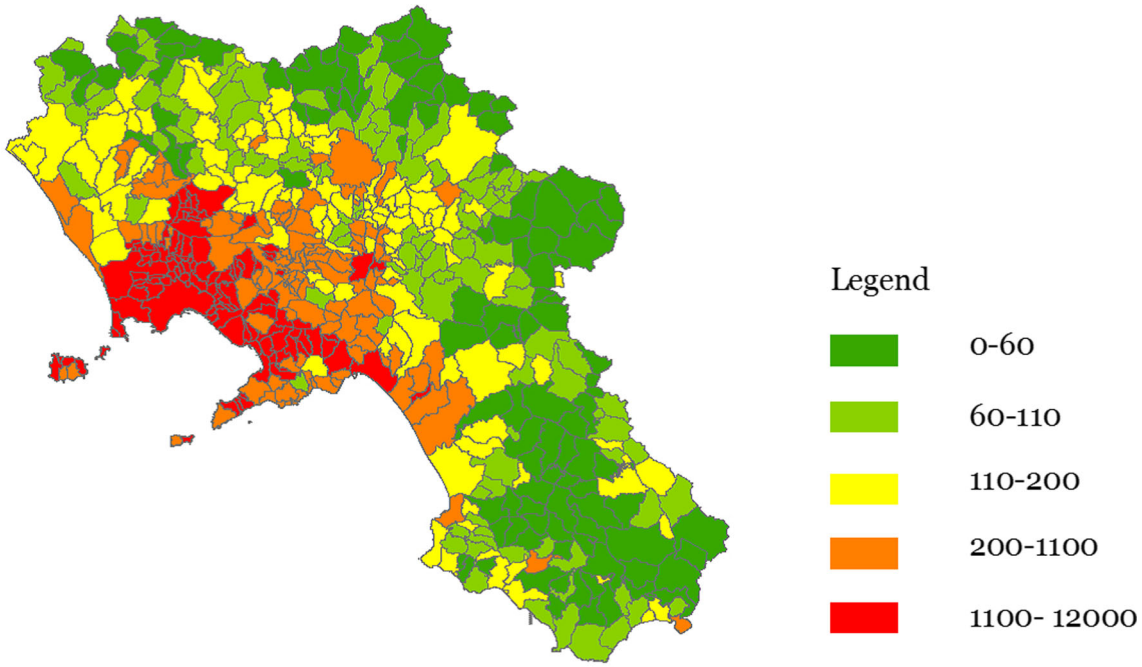

Fig. 1 Map of the population density of the Campania Region (inhabitants per Km^2)

Table 1 Labelled instances structure

Feature name	Description
*F*1	Number (*F*1): a number representing the day the data was entered. It is expressed as an integer, starting on February 24, 2020 (1) until December 21, 2020 (302)
*F*2	Hospitalized with symptoms (*F*2): the number of people that are hospitalized with symptoms
*F*3	Intensive care (*F*3): the number of people that are hospitalized in intensive care
*F*4	Total hospitalized (*F*4): the total number of people that are hospitalized also out of intensive care
*F*5	Home isolation (*F*5): the number of people that are in home isolation
*F*6	Total positives (*F*6): the total number of infected
*F*7	Variation (*F*7): the variation in the number of infected people compared to the previous day
*F*8	New positives (*F*8): the number of new people infected
*F*9	Discharged healed (*F*9): the number of people healed
*F*10	Deceased (*F*10): the number of people deceased
*F*11	Cumulative total cases (*F*11): the number of people infected since the start of the pandemic
*F*12	Molecular swabs (*F*12): the number of molecular swabs performed during the day
Forecast	New positives (*F*8) after ten days from the current date (*F*1 + 10)

by the Department of Civil Protection and made available on the website http://www.protezionecivile.gov.it/ following the official communication via a press conference at 6.00 pm by the Head of the Department of Civil Protection as extraordinary Commissioner.

The data describe in successive lines the daily situation in the Campania Region in terms of number of infected people (hospitalized, in intensive care, in home isolation, currently positives, new positives, discharged, cured, deceased, total) and swabs and cases tested.

At the time of use, the dataset included daily data from February 24, 2020 to December 31, 2020 (312 rows).

From each row, we defined a feature vector, adding a label, named Forecast, representing the new positives after ten days from the current date. The feature vector structure is reported in the Table 1.

In this way, we obtained 302 labelled instances from February 24, 2020 to December 21, 2020 (302). It is to point out that there is a negative value of new positive

Table 2 GP selected hyperparameter

Hyperparameter	Lower bound	Upper bound	Selected
Population	100	1000	100
Generations	100	500	100
Tournament size	0	100%	30%
Crossover probability	0	1	0.8
Mutation probability	0	1	0.4
Trees depth	1	10	5
Maximum number of genes	1	10	2

(− 229) in the data of June 02, 2020, which is probably a correction of the previous data. We left it unchanged.

3.1 Cross-validation and fitness measure

To build the formulae avoiding bias, we divided the dataset of Sect. 2.1 into 5 sub-sets according to the k-fold cross-validation approach (Devijver and Kittler 1982). In this way, the whole dataset was divided into 5 folds, and, in turn, one fold was used as validation set, while the remaining folds were used as training set.

As fitness measure leading GP (Affenzeller et al. 2009) we chose the minimum Root Mean Square Error (RMSE), where

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{m} \frac{(y_i - \hat{y}_i)^2}{m}} \tag{1}$$

where $\hat{y}_i$ is the prediction and y_i he true value, while m is the number of samples.

3.2 GP hyperparameters tuning

The GP experiments were made in the Matlab environment (Higham and Higham 2016).

To run GP, several hyperparameters were set, such as the population size, the maximum number of generations, the tournament type and its size, the maximum depth of trees, the maximum number of genes allowed in an individual, the permitted operators. We remark that the choice of these parameters significantly affect the final result (Sipper et al. 2018).

These choices are generally made in a manual or automatic way. In the former, the values of the hyperparameters are randomly chosen by using a trial-and-error method through an extensive series of experiments and evaluation of the corresponding performance. The latter makes use of intelligent logic able to find out the appropriate values of the hyperparameters through an iteration-based method. In this study, we used the second approach by first defining the upper and lower bounds of each hyperparameter and then choosing them by following the workflow used by the Talos library implemented for running Tensorflow-based app in Python language.[2] More specifically, we used 70% of the dataset for calibrating these parameters.

The selected hyperparameters and their ranges are reported in Table 2.

3.3 GP formulae

In the GP experiments, we were looking for formulae *f()* that would satisfy

$$\mathrm{Forecast} = f(F1, F2, \ldots, F12) \tag{2}$$

from the described data.

As aforementioned, we performed 5 main experiments, according to the fivefold cross-validation. Each experiment was repeated 100 times, and the best solution was considered. Besides, GP was applied on the whole dataset.

The resulting formulae, for each cross-validation experiment and for the whole dataset experiment, are reported in Table 3.

Table 4 shows the RMSE for each experiment, the mean value of the 5 cross-validation results and the RMSE value when the entire dataset was considered. Figure 2 graphically shows the expected and actual values of the new positives in the experiments. In particular, the graph of Exp 2 highlights the negative value of June 02, 2020 and its impact on forecasts.

Table 5 shows how the considered features are distributed among the formulae carried out in the experiments. With reference to the occurrences reported in Table 5, the most significant characteristics seem to be $F7$, $F8$, $F10$ and $F12$, i.e., the number of new positives at 10 days from the moment of observation seems strongly dependent on the current variation in the number of infected people, newly infected, deceased people and molecular swabs performed at the time of observation.

[2] https://github.com/autonomio/talos/wiki/Workflow.

Table 3 GP formulae for each experiment

Experiment	*f()*
Exp 1	Forecast = 0.4054*$F2$ − 1.62*$F1$ + 0.81*$F7$ − 0.0008*F8 − 0.81*$F10$ + 0.0016*$F12$ + 51.77
Exp 2	Forecast = 0.003128 * $F7$ − 0.01877 * $F5$ + 0.7048 * $F8$ + 0.009384 * F − 2.118 * $F10$ + 0.003128 * $F12$ + 27.18
Exp 3	Forecast = 0.8768 * $F8$ − 0.02289 * $F5$ − 9.645 * $F1$ − 0.8768 * $F10$ − 0.02289 * $F11$ + 0.005722 * $F12$ + 155
Exp 4	Forecast = 2.221 * $F1$ + 2.221 * $F3$ + 0.352 * $F7$ − 1.869 * $F10$ + 0.002304 * $F12$ + 9.558
Exp 5	Forecast = 0.2076 * $F6$ + 0.2071 * $F3$ + 0.00222 * $F4$ − 0.00222 * $F5$ − 0.0005551 * $F6$ + 0.6195 * $F7$ + 0.4136 * $F8$ + 0.0005551 * $F12$ − 112
Exp all data	Forecast = 1.948 * $F8$ + 52.29/($F8$ + 2.038) − (3.436 * (2 * $F8$ + $F9$) * (0.149 * $F8$ + 1) * (0.214 * $F8$ + 4.308))/(107) − 18.2 (7)

Table 4 The RMSE for each experiment and the mean value of the 5 cross-validation results

Experiment	RMSE
Exp1	73,31
Exp2	359.39
Exp3	88.00
Exp4	577.16
Exp5	1144.64
Mean	448,50
Exp all data	278.45

3.4 Result comparison

To compare the results, we repeated the experiments by using several algorithms widely used in the literature, that is k-Nearest Neighbors (KNN-Regression), Multi-Layer Perceptron (MLP), Support Vector Machines (SMO Regression), and Regression Tree (REPTree). All experiments were carried out by using the Waikato Environment for Knowledge Analysis (WEKA) by using the same Folds as for GP testing (Witten et al. 2016).

Table 6 shows the results. As depicted, the RMSE values are comparable with those obtained from GP, while these algorithms are not capable to provide a representation of the relationship among features involved, given their sub symbolic nature (Ilkou and Koutraki 2020).

4 Experimental results during the pandemic

In order to integrate the results into the SVIMAC-19 system, extending the forecast interval to 15 days, a new GP formula was produced with a new set of data available. We considered the Campania Region data available at the following link:https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv.

At the time of use, the dataset included daily data from February 24, 2020 to April 01, 2021 (403 rows). The considered features are the same of Table 1 except for the Forecast label, changed in the number of new positives at 15 days from the time of observation as reported in Table 7.

In this way, we obtained 388 labelled instances from February 24, 2020 (1) to March 17, 2021 (388).

The GP formula was built by using the whole dataset, and it is reported in Table 8.

Figure 3 plots real and predicted data. The RMSE achieved was 436.88 (variation explained 84.2035%).

As it can be derived by Table 8, also in this case the most significant feature is F8 which is present with a very high coefficient (3.222) in the equation, while F3 is less representative due to its medium coefficient (1.611), and lastly, F1 and F2 are very unrepresentative due to a very small coefficient (0.001135).

Then the formula has been integrated in the SVIMAC-19 system where it is still operating. The performances are valued both by the RMSE and by 4 standard measures of forecast error for both scientific and applicative fields:

Mean Error (ME), i.e., the arithmetic mean of the errors:

$$\mathrm{ME} = \frac{1}{m}\sum_{t=1}^{m} e_t \tag{3}$$

Mean Squared Error (MSE), i.e., the arithmetic mean of the squares of the errors:

$$\mathrm{MSE} = \frac{1}{m}\sum_{t=1}^{m} e_t^2 \tag{4}$$

Mean Absolute Error (MAE), i.e., the arithmetic average of the errors taken as an absolute value:

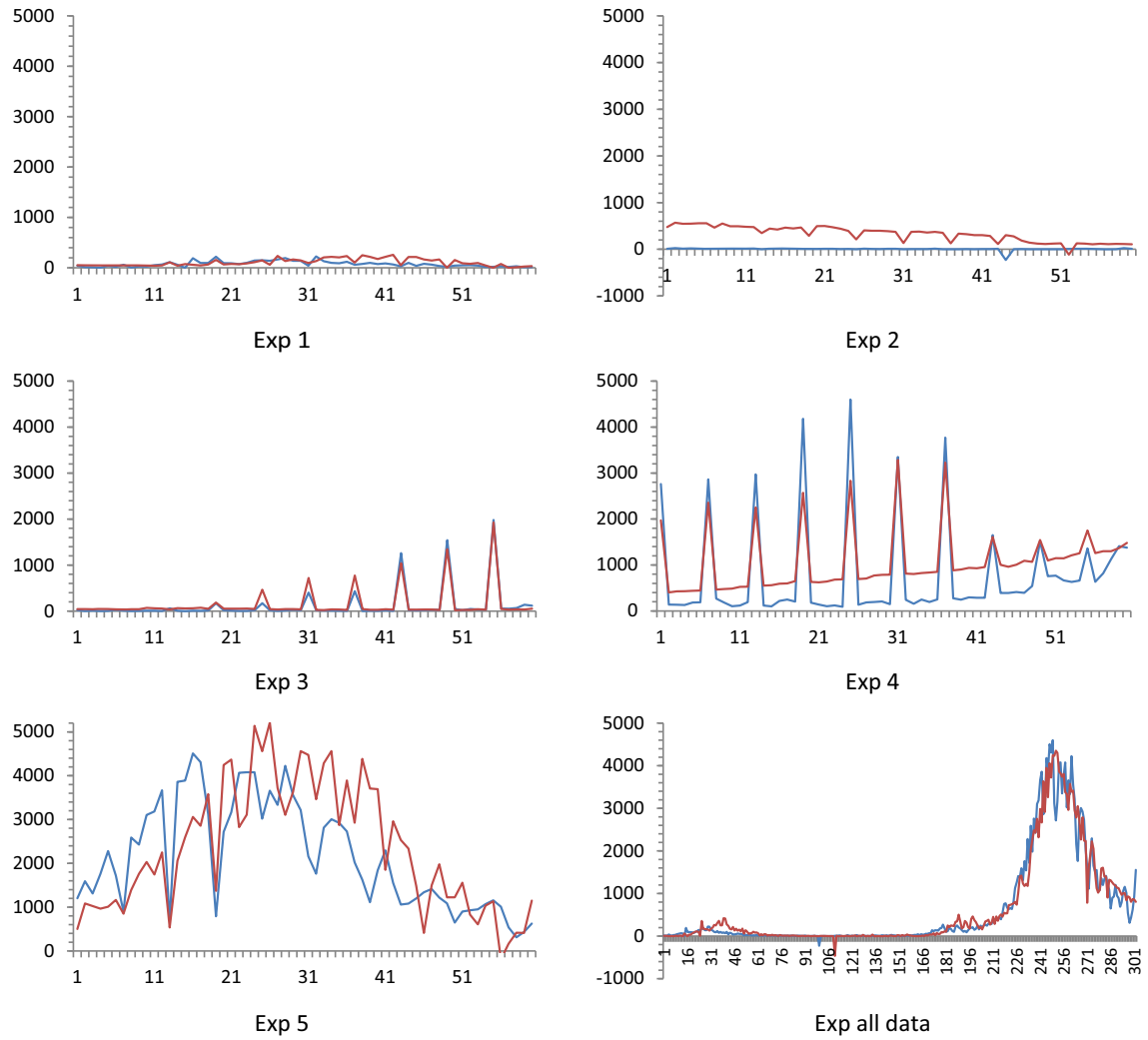

Fig. 2 Plot of predicted and real values of new positives at 10 days for each formula in Table 3. In each picture, the real values are reported in blue and the predicted values are reported in red. The graph of Exp 2 highlights the negative value of June 02, 2020

$$\text{MAE} = \frac{1}{m}\sum_{t=1}^{m}|e_t| \tag{5}$$

Mean Absolute Percentage Error (MAPE), that is the arithmetic mean of the relative percentage errors, taken as an absolute value:

$$\text{MAPE} = \frac{1}{m}\sum_{t=1}^{m}\frac{|e_t|}{y_t}100 \tag{6}$$

where y_i is the true value.

We report the experimental results during nine months of operation, i.e., from March 18, 2021 to December 18, 2021. The error measures are reported in Table 9, while the Fig. 4 reports the plot of predicted and real values.

5 Conclusions

In this paper, we used Genetic Programming to evidence dependences of the SARS-CoV-2 spread from past data in the Campania Region, in Italy. Our approach aimed to build a forecasting model by mining useful insights from the data observed over time, without taking into account any type of external information or human intervention.

Furthermore we based the prediction only from a few information, such as infected people (hospitalized, in intensive care, in home isolation, currently positives, new positives, discharged, cured, deceased, total) and swabs and cases tested.

According to our experimental results, which provide an explicit representation of relationships from the data, the number of future new positives appears to be independent from the number of people that are currently hospitalized

Table 5 Feature occurrences for each formula

Experiment	*F*1	*F*2	*F*3	*F*4	*F*5	*F*6	*F*7	*F*8	*F*9	*F*10	*F*11	*F*12
Exp 1	1	1	0	0	0	0	1	1	0	1	0	1
Exp 2	0	0	0	0	1	0	1	1	1	1	0	1
Exp 3	1	0	0	0	1	0	0	1	0	1	1	1
Exp 4	1	0	1	0	0	0	1	0	0	1	0	1
Exp 5	0	0	1	1	0	1	1	1	0	0	0	1
Exp all data	0	0	0	0	0	0	0	1	1	0	0	0
Occurrences	3	1	2	1	2	1	4	5	2	4	1	5
Occurrence%	50%	17%	33%	17%	33%	17%	67%	83%	33%	67%	17%	83%

Table 6 RMSE values for each compared method in all the experiments and the mean value of the 5 cross-validation results

Experiment	KNN-Regression	MLP	SMO Regression	REPTree
Exp1	271.75	158.24	473.01	245.12
Exp2	304.44	240.22	483.39	344.06
Exp3	364.48	311.49	444.77	345.63
Exp4	263.79	241.20	386.64	283.14
Exp5	375.88	308.67	502.30	410.00
Mean	316.07	251.96	458.02	325.59
Exp all data	313.76	169.98	444.15	214.62

Table 7 Label of the new instances

Forecast	New positives (F8) after fifteen days from the current date (F1 + 15)

Table 8 GP formula for the whole dataset

f()
Forecast = 3.222 * *F*8 − 1.611 * *F*3 + 0.001135 * *F*1 * (*F*2 − 2 * F8) + 0.001128 * *F*2 * (*F*3 − *F*8) − 6.045

with symptoms or in intensive care, and also from the number of people in home isolation, as well as from the total number of infected people since the start of the pandemic. On the contrary, the incidence of the current number of newly infected is evident.

The resulting models proved their effectiveness in predicting the number of new positives 10/15 days earlier. Then, thanks to the model adoption within a monitoring system, the experimental data were analyzed in the long term by evaluating different error measures such as Root

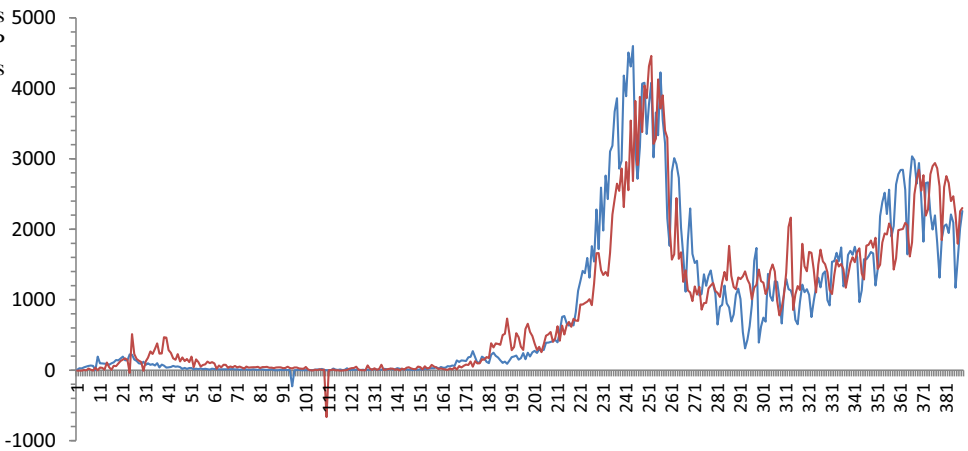

Fig. 3 Plot of predicted and real values of new positives at 15 days for the GP formula. In the picture, the real values are reported in blue and the predicted values are reported in red

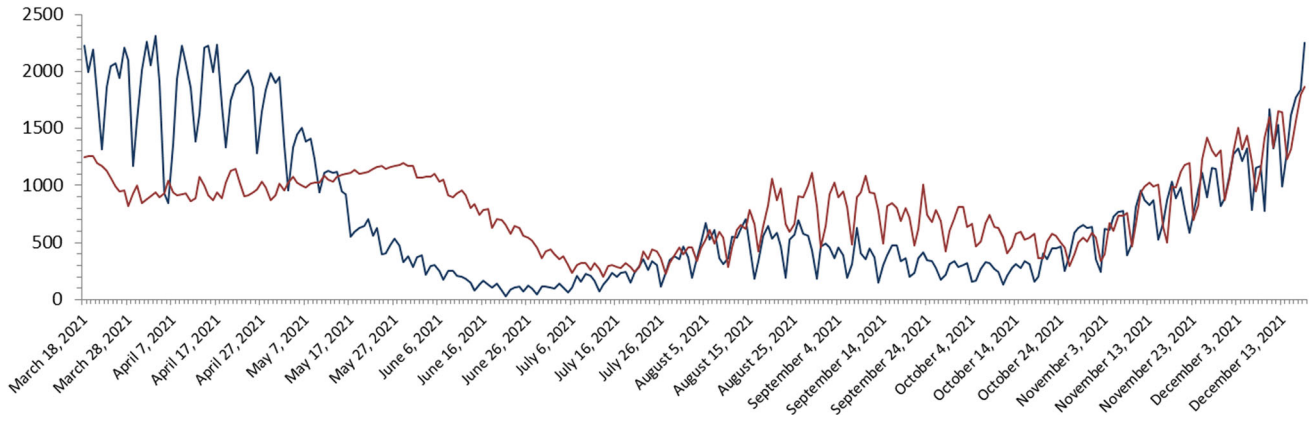

Fig. 4 Plot of predicted and real values of new positives at 15 days for the GP formula on the unseen data from March 18, 2021 to December 18, 2021 (real data are reported in blue, predicted data are reported in violet)

Table 9 Error measures as defined in (1), (3), (4), (5), (6) for the time interval from March 18, 2021 to December 18, 2021

Measure	Value
RMSE	517,98
ME	− 69,25
MSE	268,299,49
MAE	399,64
MAPE	115%

Mean Square Error, Mean Error, Mean Squared Error, Mean Absolute Error, Mean Absolute Percentage Error.

The general adherence of the forecast curve to the real trend is rather surprising. In fact, in line with the initial choices, the model has not been modified following the strengthening of the vaccination policy and the occurrence of virus mutations. This suggests that the latter have an impact mainly on the severity of the disease rather than on the spread of the virus, and this will be a topic for future work.

Acknowledgements This work has been supported by the European Union, the Italian State and the Campania Region, as part of the POR Campania FESR 2014-2020 19 (DGR n. 140 17 03 2020) in the framework of the project SVIMAC-19 Sistema Visuale Integrato di Monitoraggio e predizione Andamento Covid-19.

Author's contributions S.R. designed research and coordinated the study; G.D. performed the model set up; S.R. and G.D. analyzed and interpreted data; S.R. led the model integration into the SVIMAC-19 system; S.R. and G.D. wrote the manuscript.

Funding The authors have not disclosed any funding.

Data availability The datasets analyzed during the current study are available in the https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv repository.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent In this study, only aggregated and anonymous data were used without direct connection with specific patients.

References

Affenzeller M, Winkler S, Wagner S, Beham A (2009) Genetic algorithms, and genetic programming: modern concepts and practical applications, 1st edn. Chapman-Hall/CRC, London

Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X (2020) Artificial intelligence and machine learning to fight covid-19. Physiol Genom 52(4):200–202

Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185. https://doi.org/10.1080/00031305.1992.10475879.hdl:1813/31637

Anđelić N, Baressi ŠS, Lorencin I, Mrzljak V, Car Z (2021) Estimation of COVID-19 epidemic curves using genetic programming algorithm. Health Inform J 27(1):1460458220976728

Angluin D, Smith CH (1983) Inductive inference: theory and methods. ACM Comput Surv 15(3):237–269. https://doi.org/10.1145/356914.356918

Camporesi S, Angeli F, Dal Fabbro G (2022) Mobilization of expert knowledge and advice for the management of the Covid-19 emergency in Italy in 2020. Hum Soc Sci Commun 9:54. https://doi.org/10.1057/s41599-022-01042-6

D'Angelo G, Palmieri F (2021) Gga: a modified genetic algorithm with gradient-based local search for solving constrained optimization problems. Inf Sci 547:136–162. https://doi.org/10.1016/j.ins.2020.08.040

Devijver PA, Kittler J (1982) Pattern recognition: a statistical approach. Prentice-Hall, London

Del Giudice V, De Paola P, Del Giudice FP (2020) Covid-19 infects real estate 290 markets: short and mid-run effects on housing prices in Campania region (Italy). Soc Sci. https://doi.org/10.3390/socsci9070114

Doornik JA, Castle JL, Hendryab DF (2022) Short-term forecasting of the coronavirus pandemic. Int J Forecast 38(2):453–466

Fix E, Hodges JL (1951) Discriminatory analysis. Nonparametric Discrimination: Consistency Properties. USAF School of Aviation Medicine, Randolph Field, Texas

Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J (2020) The species Severe acute respiratory syndrome related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536–544. https://doi.org/10.1038/s41564-020-0695-z

Higham DJ, Higham NJ (2016) MATLAB Guide, 3rd edn. SIAM Society for Industrial and Applied Mathematics, Philadelphia

Katoch S, Chauhan SS, Kumar V (2020) A review on the genetic algorithm: past, present, and future. Multimed Tools Appl. https://doi.org/10.1007/s11042-020-10139-6

Kour H, Gondhi N (2020) Machine learning techniques: a survey. In: Raj J, Bashar A, Ramson S (eds) Innovative data communication technologies and application. ICIDCA 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 46, Springer, Cham. https://doi.org/10.1007/978-3-030-38040-3_31

Koza JR (1994) Genetic programming as a means for programming computers by natural selection. Stat Comput 4(2):87–112. https://doi.org/10.1007/BF00175355

Ilkou E, Koutraki M (2020) Symbolic vs sub-symbolic AI methods: friends or enemies? In: Proceedings of the CIKM 2020 workshops, October 19–20, Galway, CEUR Workshop Proceedings ISSN 1613-0073 C

Li Y, Liang M, Yin X, Liu X, Hao M, Hu Z, Wang Y, Jin L (2020) Covid-19 epidemic outside china: 34 founders and exponential growth. J Investig Med 69(1):52–55

Pak D, Langohr K, Ning J, Cortés Martínez J, Melis GG, Shen Y (2020) Modeling the coronavirus disease 2019 incubation period: impact on quarantine policy. Mathematics 8:1631. https://doi.org/10.3390/math8091631

Paterlini M (2020) On the front lines of coronavirus: the Italian response to Covid-19. BMJ 368:m1065

Rampone S, Pagliarulo C, Marena C, Orsillo A, Iannaccone M, Trionfo C, Sateriale D, Paolucci M (2021) In silico analysis of the antimicrobial activity of phytochemicals: towards a technological breakthrough. Comput Methods Programs Biomed 200:105820

Rampone S, Russo C (2012) A fuzzified brain algorithm for learning DNF from incomplete data. Electron J Appl Stat Anal 5(2):256–270

Rampone S, Valente A (2021) Evidence of the correlation between a city's air pollution and human health through soft computing. Soft Comput 25(24):15335–15343

Regione Campania (2020) Avviso per l'acquisizione di manifestazioni di interesse per la realizzazione di servizi di ricerca e sviluppo per la lotta contro il Covid-19 (DGR n. 140 del 17 marzo 2020) POR FESR Campania 2014–2020–Asse I - Misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-2019, http://www.regione.campania.it/regione/it/news/regione-informa/avviso-manifestazioni-di-interesse-per-servizi-di-ricerca-e-sviluppo-per-la-lotta-contro-il-covid-19?page=1 (Accessed Mar 2022)

Remuzzi A, Remuzzi G (2020) Covid-19 and Italy: what next? The Lancet 395(10231):1225–1228. https://doi.org/10.1016/S0140-6736(20)30627-9

Salgotra R, Gandomi M, Gandomi AH (2020) Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries. Chaos Solitons Fractals 140:110118

Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data. Science 324(5923):81–85. https://doi.org/10.1126/science.1165893

Sipper M, Fu W, Ahuja K, Moore JH (2018) Investigating the parameter space of evolutionary algorithms. Bio Data Min 11:2

Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE (2020) Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Comput Sci 1(4):197. https://doi.org/10.1007/s42979-020-00209-9

Siniscalchi S (2018) Cementificazione edilizia e paesaggi costieri. Il Caso Del Cilento Bollettino Associazione Italiana Di Cartografia 163:113–126. https://doi.org/10.13137/2282-572X/24276

Sinsbeck M, Höge M, Nowak W (2020) Exploratory-phase-free estimation of GP hyperparameters in sequential design methods-at the example of Bayesian inverse problems. Front Artif Intell 3:52. https://doi.org/10.3389/frai.2020.00052

Tu H, Tu S, Gao S, Shao A, Sheng J (2020) Current epidemiological and clinical features of COVID-19; a global perspective from China. J Infect 81:1–9. https://doi.org/10.1016/j.jinf.2020.04.011

Tuttitalia: Guida ai Comuni, alle Province ed alle Regioni d'Italia. https://www.tuttitalia.it/campania/ (2018). Accessed Mar 2020

Witten I.H., Frank E., Hall M.A., Pal C. (2016) Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann 4th Edition ISBN: 9780128042915

Zhang S, Li X, Zong M, Zhu X, Cheng D (2017) Learning k for kNN classification. ACM Trans Intell Syst Technol 8(3):1–19. https://doi.org/10.1145/2990508