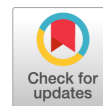


FORECASTING INFECTION FATALITY RATE OF COVID-19: MEASURING THE EFFICIENCY OF SEVERAL HYBRID MODELS



D. Seba^{a,b}, K. Belaïde^b

^a Higher School of Informatics, Sidi Bel Abbes, Algeria

^b University of Bejaia, Bejaia, Algeria

Abstract. The main goal of this paper is to delve into a crucial epidemiological metric the daily infection fatality rate in the context of the ongoing COVID-19 pandemic. The significance of understanding this metric lies in its potential to provide insights into the severity and impact of the virus on a daily basis. **Methods:** To achieve this overarching objective, we employ a comprehensive approach by applying various hybrid models that hybridize both machine learning and statistical techniques. In our pursuit of a deeper understanding, we leverage advanced machine learning algorithms, including Support Vector Machine and Random Forest. These techniques allow us to capture intricate patterns and relationships within the data, contributing to a more nuanced analysis of the infection fatality rate. The application of machine-learning models in epidemiological studies has gained prominence due to their ability to adapt to complex and evolving patterns inherent in infectious disease dynamics. Complementing our machine-learning arsenal, we integrate traditional statistical models such as ARIMA (AutoRegressive Integrated Moving Average), fractional ARIMA, and BATS (Bayesian Structural Time Series). **Results.** To assess the performance of these models, we employ key evaluation metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). These metrics serve as critical benchmarks, allowing us to quantify the accuracy and reliability of our models in predicting the daily infection fatality rate. A meticulous evaluation of model performance is crucial for ensuring the validity and of our findings. According to these measures, we see that hybrid models performed well especially ARIMA-RF model RMSE: 0.29, MSE: 0.084, MAE: 0.215 for the horizon 60 and for horizon 120 ARIMA-RF still the best performance, RMSE: 0.268, MSE: 0.071, MAE: 0.183, we get these results due to the capacity of this approach to handle complex patterns contrarily to other model ARIMA, BATS, RF and SVM. **Conclusion.** This work adopted this approach in order to build a model to predict infection fatality rate, we aspire to provide a nuanced understanding of the factors influencing the severity of the virus, ultimately contributing to the ongoing discourse on effective public health interventions and mitigation strategies.

Key words: epidemiology, infection fatality rate, forecast, statistical models, machine learning, performance, hybrid model.

ПРОГНОЗИРОВАНИЕ УРОВНЯ ЛЕТАЛЬНОСТИ ПРИ COVID-19: ОЦЕНКА ЭФФЕКТИВНОСТИ НЕКОТОРЫХ ГИБРИДНЫХ МОДЕЛЕЙ

Себа Д.^{1,2}, Белаиде К.^{1,2}

¹ Высшая школа информатики, г. Сиди-Бель-Аббес, Алжир

² Университет Беджаи, г. Беджая, Алжир

Резюме. История вопроса. Основная цель этой статьи заключается в углублении понимания роли важнейшего эпидемиологического показателя — коэффициента летальности при заражении на фоне продолжающейся пандемии COVID-19. Важность понимания этого показателя состоит в том, что он позволяет ежедневно

Адрес для переписки:

Джиллали Себа
22016, Алжир, г. Сиди Бель Аббес, ул. 8 мая 1945 года,
Высшая школа информатики.
Тел.: +213-7-82-01-74-34. E-mail: d.seba@esi-sba.dz

Contacts:

Djillali Seba
22016, Sidi Bel Abbes, Algérie, Rue du 8 Mai 1945,
École Supérieure en Informatique.
Phone: +213-7-82-01-74-34. E-mail: d.seba@esi-sba.dz

Для цитирования:

Себа Д., Белаиде К. Прогнозирование уровня летальности при COVID-19: оценка эффективности некоторых гибридных моделей // Инфекция и иммунитет. 2024. Т. 14, № 2. С. 313–319. doi: 10.15789/2220-7619-FIF-17548

Citation:

Seba D., Belaïde K. Forecasting infection fatality rate of COVID-19: measuring the efficiency of several hybrid models // Russian Journal of Infection and Immunity = Infektsiya i immunitet, 2024, vol. 14, no. 2, pp. 313–319. doi: 10.15789/2220-7619-FIF-17548

© Seba D., Belaïde K., 2024

DOI: <http://dx.doi.org/10.15789/2220-7619-FIF-17548>

оценивать выраженность вирусного инфицирования и ее воздействие. *Материалы и методы.* Для достижения этой всеобъемлющей цели мы применяем комплексный подход, используя различные гибридные модели, объединяющие машинное обучение и статистические методы. Углубленное понимание значимости указанного параметра достигается с использованием передовых алгоритмов машинного обучения, включая машину опорных векторов и случайные деревья решений. Эти методы позволяют обнаруживать сложные закономерности и взаимосвязи в данных, способствуя более детальному анализу уровня смертности от инфекций. Применение моделей машинного обучения в эпидемиологических исследованиях приобрело известность благодаря их способности адаптироваться к сложным и развивающимся закономерностям, присущим динамике инфекционных заболеваний. Дополняя доступный арсенал машинного обучения, мы совмещаем традиционные статистические модели, такие как ARIMA (авторегрессионное интегрированное скользящее среднее), дробное ARIMA и BATS (байесовский структурный временной ряд). Эти модели обеспечивают проверенную временем и строгую статистическую основу для проводимого анализа, позволяя выявить временные зависимости и тенденции коэффициента летальности при заражении. Синергия машинного обучения и статистических моделей дает нашему исследованию целостную перспективу, обеспечивая надежное и всестороннее исследование эпидемиологической ситуации. *Результаты.* Для оценки эффективности этих моделей мы используем ключевые показатели оценки, включая среднеквадратическую ошибку (RMSE), среднеквадратическую ошибку (MSE) и среднюю абсолютную ошибку (MAE). Эти показатели служат важнейшими ориентирами, позволяя нам количественно оценить точность и надежность наших моделей при прогнозировании ежедневного коэффициента летальности при заражении. Тщательная оценка эффективности модели имеет решающее значение для обеспечения достоверности наших выводов. Согласно этим измерениям, гибридные модели показали хорошие результаты, особенно модель ARIMA-RF RMSE: 0.29, MSE: 0.084, MAE: 0.215 для горизонта 60 и для горизонта 120 ARIMA-RF по-прежнему демонстрировали лучшую производительность, RMSE: 0.268, MSE: 0.071, MAE: 0.183 позволили получить такие результаты благодаря способности этого подхода обрабатывать сложные шаблоны в отличие от моделей ARIMA, BATS, RF и SVM. *Заключение.* В данной работе применялся подход для построения модели по прогнозированию уровня коэффициента летальности при заражении, нацеленный на предоставление детального понимания факторов, влияющих на тяжесть вирусного инфицирования. В конечном итоге это будет способствовать продолжающемуся обсуждению эффективных мер общественного здравоохранения и стратегий нивелирования последствий.

Ключевые слова: эпидемиология, коэффициент летальности при заражении, прогноз, статистические модели, машинное обучение, производительность, гибридная модель.

Introduction

A highly contagious respiratory illness, COVID-19 caused by the SARS-CoV-2 virus was initially discovered in China, in December 2019, and since then, it has spread over the world.

This pandemic has had significant impacts on many aspects of life, including public health, the economy, education, and social interactions. For this reason, forecasting COVID-19 is an important tool in managing the pandemic, helping to minimize its impact and inform public health interventions.

Numerous researchers extensively explored this pandemic. For instance Alzahrani et al. [4] used ARIMA model to predict the spread of the pandemic, Dahesh et al. [8] treated the new cases using ARIMA model, Roy et al. [16] focused on spatial prediction. Note-worthy contributions also include the research conducted by Rath et al. [15], Chen [9], Lukman et al. [13], Yousaf et al. [23].

Numerous researchers have examined this phenomena through the application of genetic algorithm, such as the works of Deif et al. [10], Salgorta et al. [17], Acosta et al. [1].

The deep learning tools are also used to predict the new cases such as the work of Alazab et al. [3], Tamang et al. [19] Kapoor et al. [12] and Namasudra et al. [14] use neural networks, Zeroual et al. [24] make

a comparative study between different deep learning models, ArunKumar et al. [5] compared between statistical models ARIMA, seasonal ARIMA model and machine learning models Gated Recurrent unit (GRU), Long-Short term memory (LSTM).

IFR stands for infection fatality rate, which is the proportion of people who die from an infectious disease among all those who have been infected, was also estimated and forecasted in many works such as Singh et al. [18], Vattay et al. [22] Forecast the outcome and estimating the epidemic model parameters from the fatality time series. Ahmar et al. [2] use ARIMA and nonlinear AR model.

In the remainder of this paper we deal with forecasting daily IFR using hybrid models then we evaluate their effectiveness. In the second section we present both the data and the descriptive statistics, which provide insights into the behavior of the phenomenon. Moving on to the third section we elaborate the methodology employed in our study, The final section encompasses the implementation of our approach which includes statistical models (ARIMA, BATS) and hybrid models (ARIMA-SVM, BATS-SVM, BATS-RVM, ARIMA-Random Forest and BATS-Random Forest), These models are subsequently subjected to comparison using performance metrics such as RMSE, MSE, and MAE.

Materials and methods

Forecasting Models

ARIMA and ARFIMA model

Autoregressive integrated moving average (ARIMA) models predict future values based on past values, it gauges the strength of one dependent variable relative to other changing variables.

A stochastic process $(X_t)_{t \geq 0}$ is said to be an ARIMA(p, d, q) an integrated mixture autoregressive moving average model if it satisfies the following equation:

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad \forall t \geq 0 \quad (1)$$

where $d \in \mathbb{N}$, L is lag operator, $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d. errors, with $\sigma^2 < \infty$.

$$\phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p) \text{ with } \phi_p \neq 0$$

$$\theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q) \text{ with } \theta_q \neq 0$$

– In the case of $d = 0$, we obtain ARMA(p, q) process;

– In the case of $d \in \mathbb{R}$, coincide with Fractional ARIMA(p, d, q) process.

BATS model

The BATS (Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components) model is a time series forecasting model that was proposed by De Livera et al. [11].

Box-Cox Transformation component is used to transform the data to achieve normality and stabilize the variance. The ARMA (Autoregressive Moving Average) Errors component is used to model the residuals of the time series data, which are assumed to be independent and identically distributed. Finally, the Seasonal component is used to model the seasonal patterns in the data.

Random Forest (RF)

The random forest for regression algorithm is a machine learning algorithm that combines multiple decision trees to predict continuous target variables. The algorithm works as follows: Select a random subset of the training data, with replacement.

Construct a decision tree for the subset of data by recursively partitioning the data into subsets based on the values of the input features. At each node, randomly select a subset of features to consider for splitting.

Repeat the previous steps to create multiple decision trees.

For prediction, pass the input data through all the decision trees and obtain the predicted target variable for each tree.

Aggregate the predictions of all trees to obtain the final prediction. This can be done by taking the average of the predicted values or using weighted averaging.

SVM model

Support vector machine (SVM) analysis is a popular machine learning tool for classification and regression, it is considered a nonparametric technique because it relies on kernel functions.

Given deviation data of training $x_i (i = 1, 2, \dots, m)$ where $x_i \in \mathbb{R}^n$ is the input vector with n -dimension, $y_i \in \mathbb{R}$ is the associated desired output value of x_i . Then the SVMs model is formulated as follows:

$$f(x) = w\phi(x) + b \quad (2)$$

Where $\phi(x)$ is called the feature that is non linearly mapping from the input space x . The w and b are coefficients that are estimated by minimizing the regularized risk function shown in formula:

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^m L(y - f(x), x) \quad (3)$$

where C is the regularized constant determining the trade-off between the empirical error and the regularization term. The larger the constant C is, the more the minimum experience risk is emphasized, and the lower the generalization of function f .

Using the Lagrange function and duality theory, and with the kernel function $k(x, x)$ introduced, the function given in 3 can be transformed into a quadratic programming problem as follows:

$$\begin{aligned} \min \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j) + \\ + \epsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \\ 0 \leq \alpha_i, \alpha_i^* \leq C, \end{aligned} \quad (4)$$

where α_i and α_i^* are Lagrange multipliers. They are obtained by solving this quadratic programming problem, and the input vector x_i corresponding the nonzero α_i and α_i^* is the support vector. Thus, we transform 6 the following equation:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^m (\alpha_i + \alpha_i^*) k(x_i, x_j) + b \quad (5)$$

$$i = 1, 2, \dots, m$$

Empirical results and discussion

Infection fatality rate (IFR) is a measure used to assess the proportion of infected individuals with fatal outcomes. Here is the formula used to calculate daily IFR for COVID-19:

$$\frac{\text{Number of daily new confirmed deaths}}{\text{Number of daily new confirmed cases}} \times 100 \quad (6)$$

Source of data: *World health organization*.

We have dealt with new cases and new deaths over the world from January, 3rd 2020 to March 16th 2023 Using (6), we calculate IFR index.

As shown in Fig. 1 the IFR is may appear higher in the first three months but it subsequently decreased significantly. This is due to the social awareness and

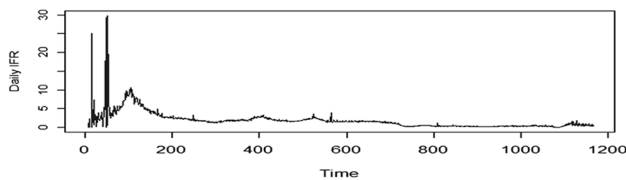


Figure 1. Daily IFR of COVID-19 from January, 3rd 2020 to March 16th 2023

the effective implementation of public health measures, such as mask mandates, social distancing, and lockdowns.

Two years into the pandemic, the IFR dropped significantly, approaching close to zero, largely due to widespread vaccination among the population, especially those at higher risk of severe illness, the overall death rate can decrease.

In the early stages of the pandemic, testing availability was limited, and many mild or asymptomatic cases went unreported. As testing capacity increases and more people get tested, health authorities can identify a larger proportion of mild cases.

Methodology

The methodology used in our work is to implement a statistical approach to model the data.

$$X_t = Y_t + U_t \tag{7}$$

X_t is decomposed into two parts linear component Y_t and nonlinear component U_t , we apply statistical models (ARIMA and BATS) which are more suited to linear pattern $\hat{Y}(t)$, then the difference $\varepsilon_t = X_t - \hat{Y}(t)$ is the residual.

The residual series ε_t contains the nonlinear parts, thus we use SVM and RVM to fit residuals gives the predicted values $\hat{U}(t)$, finally we combine the two predictive results to get:

$$\hat{X}(t)_t = \hat{Y}(t)_t + \hat{U}(t) \tag{8}$$

Results

We decompose the data into training data and test data, we use test data as horizon of forecasting to validate the results. We treat two cases, in the first case we use horizon of 60 days (short term) in other words the last 60 observations are test data. the second case horizon is 120 days (long term).

In Table 1 We provide a summary of how our data is described, we detect the missing values (Na’s). In this case the mean is close to the median it suggests that the data has a relatively symmetrical distribution, which can be a useful insight for understanding the central tendency and overall shape of the dataset.

Table 1. Descriptive statistics for Daily IFR

min	1stQu	Median	Mean	3rdQu	max	NA’s
0	0.373 9	1.5395	1.817 1	2.167 6	29.867 7	8

From Table 2 we can conclude some results about our data:

The primary purpose of the KPSS test is to check for the presence of a unit root in the time series data. A unit root indicates nonstationarity. If the test statistic is greater than the critical values at a chosen significance level, you fail to reject the null hypothesis, suggesting that the data is nonstationary and this is our case $0.01 < 0.05$.

Identifying nonstationarity is important because many time series forecasting models assume stationarity, and addressing nonstationarity may involve transformations or differencing to make the data suitable for modeling.

Kolmogorov–Smirnov test is a valuable statistical tool for assessing goodness of fit between a sample distribution and a theoretical distribution (Normal distribution in our case).

he primary utility of the Terasvirta test is to detect nonlinearities in time series data, instead of using linear models, you may choose to employ nonlinear modeling techniques and that is explain our choice to hybridize statistical models and machine learning models which handle nonlinear patterns.

The Hurst exponent serves as a valuable tool for evaluating whether long memory models are suitable candidates and for estimating the memory parameter in an ARFIMA model.

$d = H - 1/2$, thus $d = 0.49908$ which is close to 0.5, This proximity to 0.5 implies that ARFIMA model non-invertible suggesting that short memory process such as ARIMA more suitable to model our data.

From Figure 2 the ACF function did not decrease hyperbolically which confirm that short term memory model is a good candidate, and we remark clearly from the PACF function has a periodic component.

Handling missing values

We remove missing values (NA’s), it appears when the number of the new cases is 0, thus it is more accurate to replace missing values with zero’s.

ARIMA:

Due to the ARIMA model’s efficiency in analyzing time series, we first apply it. By using the Box-Jenkins approach and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to select the the best ARIMA model, which is in our case ARIMA (5,1,3), we obtain the following equation.

For the horizon of forecasting $h = 60, 120$.

$$(1+0.2104L-0.1261L^2-0.0137L^3-0.3397L^4-0.3157L^5)X_t = (1+0.7631L-0.0849L^2+0.269L^3)\varepsilon_t \tag{9}$$

We have used ARFIMA function to fit the model, we get a Fractional AutoRegressive model FAR(1,0.4998,2) the integrated fraction d is close to 0.5 which make the model not invertible.

We have used Maximum likelihood method to estimate the parameters of ARIMA and ARFIMA models.

BATS:

For the horizon $h = 60, 120$.

We perform BATS model, the output in R language is BATS (1, 1,4, -, -). The first is the ω parameter of the Box-Cox transform, the second is ARMA order of the errors, the third is ϕ trend damping, the fourth is the seasonal periods, which in your case are none. $\omega = 1$, meaning that indeed, there is no Box-Cox transformation.

ARMA order of errors is ARMA(1,4) with autoregressive coefficients 0.608 and moving average coefficients 0.658, 0.415, 0.369635, 0.52285 for the horizon $h = 120$.

SVM:

For the horizon $h = 60$.

SVM models have various hyperparameters that can be tuned using cross validation method, we have used epsilon regression which is tolerance error, we use also radial kernel $K(x_j, x_k) \exp(x_j x_k^2)$ it allows to capture nonlinear patterns. $\epsilon = 0.2$, gamma = 2, cost = 512.

For the horizon $h = 120$.

The best hyperparameters $\epsilon = 0.2$, cost = 4, gamma = 1

Random Forest:

For the horizon $h = 60$.

As Random forest is an ensemble of decision trees, thus the number of trees is 500 and mean of squared residuals is 3.001598.

For the horizon $h = 120$.

The number of trees is 500 and mean of squared residuals is 3.162188.

Forecasting Hybrid models

We treat the residuals of ARIMA with SVM then with Random Forest model in order to improve the forecasting results, we do the same thing with BATS-SVM and BAATS RVM.

We tune the hyperparameters using cross validation technique.

Table 2. Some characteristics of daily Infection Fatality Rate

Tests	Daily IFR	Comment
KPSS test	0.01	Non stationary
Kolmogorov-Smirnov Test	2.2e-16	Non normal
Terasvirta test	2.2e-16	Non linear
Hurst Exponent	0.99908	Short term memory

Illustration

Performance Measures

RMSE, MAE, and MSE are commonly used evaluation metrics in machine learning and statistics to assess the performance of models. They are used to measure the accuracy of predicted continuous values compared to the actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - z_i)^2} \tag{10}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - z_i)^2 \tag{11}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - z_i|) \tag{12}$$

Discussion

A small MAE means that our model is excellent at predictions, while a large MAE suggests that our model does not perform well at predictions. Unlike MSE, we do not square the residuals, thus MAE is more robust to outliers.

A higher MSE indicates that the model will be penalized for making predictions that significantly differ from the actual value. This means that a large difference between predicted and actual values will be more heavily penalized in MSE than in MAE. RMSE and MSE are sensitive to outliers.

We remark in Fig. 3 and 4 (see cover III) Hybrid models in two cases for $h = 60$ and $h = 120$ perform better than ARIMA and BATS because these models are linear and can not handle nonlinear pattern, contrarily SVM can treat nonlinear patterns due to its form of kernel, Random Forest has a good performance because it aggregate the results of many decision tree.

We are unable to make predictions using ARFIMA due to the parameter d being extremely close to 0.5 which makes the model non-invertible The advantage

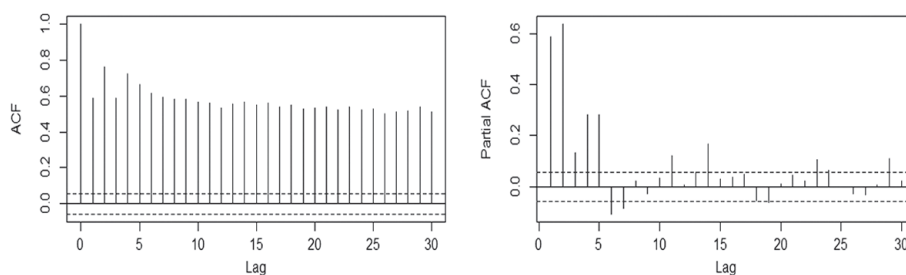


Figure 2. ACF and PACF for Daily IFR

Table 3. Performance Measures

h = 60	RMSE	MSE	MAE	h = 120	RMSE	MSE	MAE
ARIMA	0.367	0.135	0.248	ARIMA	0.369	0.136	0.272
ARFIMA	NaN	NaN	NaN	ARFIMA	NaN	NaN	NaN
BATS	0.388	0.151	0.265	BATS	0.363	0.132	0.289
SVM	0.362	0.131	0.243	SVM	0.345	0.129	0.266
RF	0.308	0.095	0.234	RF	0.377	0.142	0.278
BATS-SVM	0.345	0.119	0.251	BATS-SVM	0.345	0.129	0.264
BATS-RF	0.301	0.094	0.219	BATS-RF	0.360	0.094	0.289
ARIMA-SVM	0.324	0.105	0.216	ARIMA-SVM	0.342	0.117	0.266
ARIMA-RF	0.290	0.084	0.215	ARIMA-RF	0.268	0.071	0.183

of hybrid models is that we can deal with linear components (ARIMA and BATS) and nonlinear component (SVM and RF).

In Fig. 3 (see cover III), for $h = 120$, BATS has a good performance then ARIMA because it deals with short term memory phenomenon, same thing for BATS but this model can deal with complex pattern in time series such seasonal component and it assumes that there is a correlation between the errors.

In Fig. 4 (see cover III), for $h = 120$, ARIMA-RF has a good performance for the 60 days, but thereafter, its performance declines, primarily attributed to the limited memory capacity within Random Forest (RF) models.

Based on the performance measures in Table 3, it is evident that ARIMA-RF consistently demonstrates the superior performance in both scenarios.

Conclusion

To sum up, we have studied the behavior of daily IFR of COVID-19 using statistical models and machine learning models, we have fitted the values of daily IFR which help us to understand the phenomena. To improve the forecasting results we have put into practice hybrid models such as ARIMA-SVM, BATS-SVM, ARIMA-RF and BATS-RF, basing on performance measures ARIMA-RF is the best model.

Acknowledgement

We acknowledge the support of “Direction Générale de la Recherche Scientifique et du Développement Technologique DGRSDT”.MESRS ALGERIA.

References

- Acosta-González E., Andrada-Félix J., Fernández-Rodríguez F. On the evolution of the COVID-19 epidemiological parameters using only the series of deceased. A study of the Spanish outbreak using Genetic Algorithms. *Math. Comput. Simul.*, 2022, vol. 197, pp. 91–104. doi: 10.1016/j.matcom.2022.02.007
- Ahmar A.S., Boj E. Application of neural network time series (Nnar) andarima to forecast infection fatality rate (ifr) of COVID-19 in Brazil. *International Journal on Informatics Visualization*, 2021, vol. 5, no. 1, pp. 8–10. doi: 10.30630/ijov.5.1.372
- Alazab M., Awajan A., Mesleh A., Abraham A., Jatana V., Alhyari S. COVID-19 prediction and detection using deep learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 2020, no. 12, pp. 168–181.
- Alzahrani S.I., Aljamaan I.A., Al-Fakih E.A. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J. Infect. Public Health*, 2020, vol. 13, no. 7, pp. 914–919. doi: 10.1016/j.jiph.2020.06.001
- ArunKumar K.E., Kalaga D.V., Kumar C.M.S., Kawaji M., Brenza T.M. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal*, 2022, vol. 61, no. 10, pp. 7585–7603. doi: 10.1016/j.aej.2022.01.011
- Beran J. Statistics for long- memory processes. *CRC press*, 1994. Vol. 61. doi: 10.1201/9780203738481
- Box G.E., Jenkins G.M., Reinsel G.C., Ljung G.M. Time series analysis: forecasting and control; 5th ed. *John Wiley and Sons*, 2015. doi: 10.1002/9781118619193
- Dehesh T., Mardani-Fard H.A., Dehesh P. Forecasting of COVID-19 confirmed cases in different countries with ARIMA models. *MedRxiv*, 2020.03.13.20035345. doi: 10.1101/2020.03.13.20035345
- Chen J.M. Novel statistics predict the COVID-19 pandemic could terminate in 2022. *J. Med. Virol.*, 2022, vol. 94, no. 6, pp. 2845–2848. doi: 10.1002/jmv.27661
- Deif M.A., Solyman A.A., Hammam R.E. ARIMA model estimation based on genetic algorithm for COVID-19 mortality rates. *International Journal of Information Technology and Decision Making*, 2021, vol. 20, no. 6, pp. 1775–1798. doi: 10.1142/S0219622021500528
- De Livera A.M., Hyndman R.J., Snyder R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 2011, vol. 106, no. 496, pp. 1513–1527. doi: 10.1198/jasa.2011.tm09771
- Kapoor A., Ben X., Liu L., Perozzi B., Barnes M., Blais M., O'Banion S. Examining COVID-19 forecasting using spatio-temporal graph neural networks. *arXiv*, 2020: 2007.03113 [Preprint]. doi: 10.48550/arXiv.2007.03113
- Lukman A.F., Rauf R.I., Abiodun O., Oludoun O., Ayinde K., Ogundokun R.O. COVID-19 prevalence estimation: Four most affected African countries. *Infect. Dis. Model.*, 2020, vol. 5, pp. 827–838. doi: 10.1016/j.idm.2020.10.002
- Namasudra S., Dhamodharavadhani S., Rathipriya R. Nonlinear Neural Network Based Forecasting Model for Predicting COVID-19 Cases. *Neural. Process. Lett.*, 2023, vol. 55, no. 1, pp. 171–191. doi: 10.1007/s11063-021-10495-w

15. Rath S., Tripathy A., Tripathy A.R. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes Metab. Syndr.*, 2020, vol. 14, no. 5, pp. 1467–1474. doi: 10.1016/j.dsx.2020.07.045
16. Roy S., Bhunia G.S., Shit P.K. Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. *Model. Earth Syst. Environ.*, 2021, vol. 7, no. 2, pp. 1385–1391. doi: 10.1007/s40808-020-00890-y
17. Salgotra R., Gandomi M., Gandomi A.H. Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming. *Chaos Solitons Fractals*, 2020, no. 138: 109945. doi: 10.1016/j.chaos.2020.109945
18. Singh A., Bajpai M.K. SEIHCARD Model for COVID-19 spread scenarios, disease predictions and estimates the basic reproduction number, case fatality rate, hospital, and ICU beds requirement. *Computer Modeling in Engineering & Sciences*, 2020, vol. 125, no. 3, pp. 991–1031. doi: 10.32604/cmescs.2020.012503
19. Tamang S.K., Singh P.D., Datta B. Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique. *Global Journal of Environmental Science and Management*, 2020, vol. 6, special iss. (COVID-19), pp. 53–64. doi: 10.22034/GJESM.2019.06.SI.06
20. Tipping M. The relevance vector machine. *Advances in Neural Information Processing Systems 12*. 1999, pp. 652–658.
21. Vapnik V. The Nature of Statistical Learning Theory. *Springer, New York*, 1995.
22. Vattay G. Forecasting the outcome and estimating the epidemic model parameters from the fatality time series in COVID-19 outbreaks. *Phys. Biol.*, 2020, vol. 17, no. 6: 065002. doi: 10.1088/1478-3975/abac69
23. Yousaf M., Zahir S., Riaz M., Hussain S.M., Shah K. Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos Solitons Fractals*, 2020, no. 138: 109926. doi: 10.1016/j.chaos.2020.109926
24. Zeroual A., Harrou F., Dairi A., Sun Y. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos Solitons Fractals*, 2020, no. 140: 110121. doi: 10.1016/j.chaos.2020.110121

Авторы:

Себа Д., д.мат.н., доцент, лаборатория прикладной математики, факультет математики, Высшая школа информатики, г. Сиди-Бель-Аббес, Алжир;

Белаиде К., д.мат.н., профессор, лаборатория прикладной математики, факультет математики, Университет Беджаи, г. Беджая, Алжир.

Authors:

Seba D., Doctor in Mathematics, Assistant Professor, Laboratory of Applied Mathematics, Department of Mathematics, Higher School of Computer Science (Ecole Supérieure en Informatique), Sidi-Belabbes, Algeria;

Belaïde K., Doctor in Mathematics, Full Professor, Laboratory of Applied Mathematics, Department of Mathematics, University of Bejaia, Bejaia, Algeria.

Поступила в редакцию 08.12.2023
Отправлена на доработку 06.04.2024
Принята к печати 07.04.2024

Received 08.12.2023
Revision received 06.04.2024
Accepted 07.04.2024

Иллюстрация к статье «Прогнозирование уровня летальности при COVID-19: оценка эффективности некоторых гибридных моделей» (авторы: Д. Себа, К. Белаиде) (с. 313–319)

Illustration for the article “Forecasting infection fatality rate of COVID-19: measuring the efficiency of several hybrid models” (authors: Seba D., Belaide K.) (pp. 313–319)

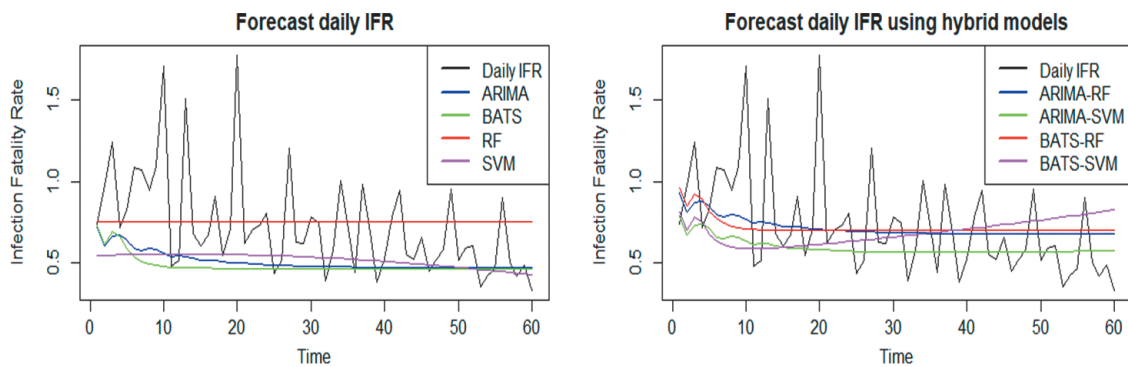


Figure 3. Daily IFR Forecasting with horizon $h = 60$

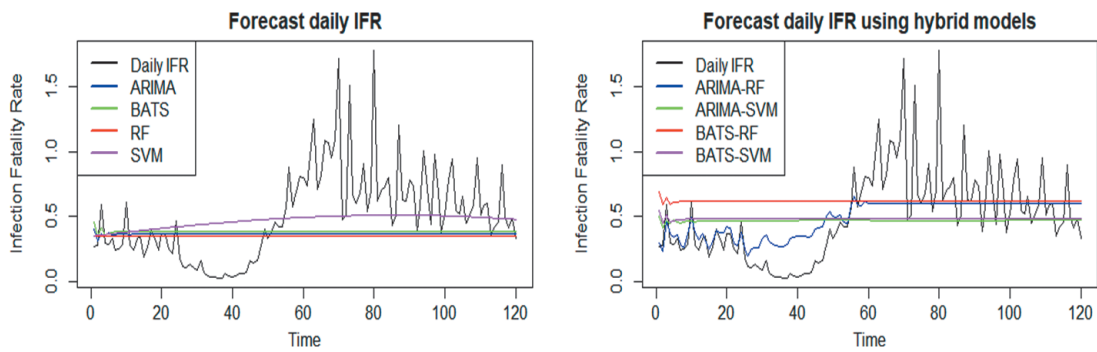


Figure 4. Daily IFR Forecasting with horizon $h = 120$