DOI: 10.15514/ISPRAS-2024-36(6)-13

Time Series Models using in Prediction of COVID-19 Infection Cases in Mexico

K. V. Cortés-Martínez, ORCID: 0000-0002-5578-7125 <keila.cortes.mtz@gmail.com>
H. Estrada-Esquivel, ORCID: 0000-0002-1466-7581 <hugo.ee@cenidet.tecnm.mx>
A. Martínez-Rebollar, ORCID: 0000-0002-1071-8599 <alicia.mr@cenidet.tecnm.mx>

TECNM/Centro Nacional de Investigación y Desarrollo Tecnológico,
Cuernavaca, Morelos, México.

Abstract. The COVID-19 pandemic was the first health crisis to affect the entire world in this century. The data captured revealed a lack of organization and control in health measures, containment, and mitigation policies, as well as a lack of planning and coordination in the use of medical supplies, which motivated the development of prediction models that provided predictive information on the evolution of the pandemic. In this work, a time series of accumulated cases of infection was generated through official data provided by the Ministry of Health of the Government of Mexico. Six deterministic and stochastic predictive models were applied to this information to compare their efficiency in predicting cases of COVID-19 infection. These models were applied to data from two cities in Mexico, Colima and the State of Mexico. The study concludes that the ARIMA and ANN MLP models adapt better to the data that is generated daily, therefore, they have an improved prediction capacity.

Keywords: COVID-19; Machine learning; Predictive models; ARIMA; ANN MLP.

For citation: Cortés-Martínez K. V., Estrada-Esquivel H., Martínez-Rebollar A. Time series models using in prediction of COVID-19 infection cases in Mexico. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 231-246. DOI: 10.15514/ISPRAS-2024-36(6)-13.

Прогнозирование случаев заражения COVID-19 в Мексике на основе моделей временных рядов

К. В. Кортес-Мартинес, ORCID: 0000-0002-5578-7125 <keila.cortes.mtz@gmail.com>
Х. Эстрада-Эскивел, ORCID: 0000-0002-1466-7581 <hugo.ee@cenidet.tecnm.mx>
А. Мартинес-Ребояр, ORCID: 0000-0002-1071-8599 <alicia.mr@cenidet.tecnm.mx>

Национальный центр исследований и технологических разработок (TECNM), Куэрнавака, Морелос, Мексика.

Аннотация. Пандемия COVID-19 стала первым кризисом в области здравоохранения, затронувшим весь мир в этом столетии. Собранные данные выявили отсутствие организации и контроля в мерах здравоохранения, сдерживании и смягчении последствий, а также отсутствие планирования и координации в использовании предметов медицинского назначения, что побудило к разработке моделей прогнозирования, которые предоставили прогнозную информацию о развитии пандемии. В этой работе временные ряды накопленных случаев заражения были получены с помощью официальных данных, предоставленных Министерством здравоохранения правительства Мексики. К этой информации были применены шесть детерминированных и стохастических прогностических моделей для сравнения их эффективности в прогнозировании случаев заражения COVID-19. Эти модели были применены к данным из двух городов Мексики, Колимы и штата Мексика. В исследовании делается вывод о том, что модели ARIMA и ANN MLP лучше адаптируются к данным, которые генерируются ежедневно, поэтому они имеют улучшенную способность прогнозирования.

Ключевые слова: пандемия COVID-19; машинное обучение; прогнозирующие модели; модель ARIMA; модель ANN MLP.

Для цитирования: Кортес-Мартинес К. В., Эстрада-Эскивел Х., Мартинес-Ребояр А. Прогнозирование случаев заражения COVID-19 в Мексике на основе моделей временных рядов. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 231–246 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–13.

1. Introduction

The COVID-19 pandemic emerged in December 2019 in the city of Wuhan, China; This event set a precedent in the study of diseases and public health emergency declarations due to its rapid spread globally [1]. In Mexico, the first case was recorded in February 2020 [2], despite containment measures [3], increases in the number of cases were observed, generating multiple waves of contagion until 2021. At this point, various models were created to predict the evolution of the pandemic using time series, but their limitations are still recognized, due to the variability of the data, the appearance of new variants of the virus, other socioeconomic factors, and the effectiveness of various control measures. Therefore, the objective of this paper is to contrast the efficiency of different models used in the literature for the analysis in predicting infection cases, considering only the data collected by the Ministry of Public Health in Mexico [4], through the use of time series. The rest of this work is organized as follows: In section 2, a summary of related works is presented. Section 3 details the methodology proposed in this article, from obtaining and analyzing the data set to the representation and experimentation with the various predictive models found in the literature. Section 4 shows the results and discussion. Finally, section 5 presents the conclusions and future work.

2. Related work on time series prediction

The data that describes the evolution of infections were used in several research groups for implementing predictive models and also to analyze, both statistically [4] and focused on machine learning [5-6], different characteristics of the COVID-19 pandemic, highlighting its effectiveness in different contexts, for example:

- *Prediction of outbreak trends:* Linear prediction models, support vector machines and exponential smoothing were used in the research of [7] and [8].
- *Infection wave prediction:* [9] and [10] applied LSTM and RNN to describe the fluctuations in the increase in the number of infection cases.
- *Prediction of positive cases:* [11] used an MLP to predict the maximum number of positive patients, [12] [13] demonstrated the effectiveness of the ARIMA and Prophet model, [14] [15] used a model with LSTM-GRU to the same task, however this latest work added the calculation of the future transmission of the virus.

In the current debate about which predictive model can efficiently provide, plan and address response strategies and resource allocation, the limitations of each of them are identified, mainly due to the variability and quality of the data that feed the databases, the emergence of new virus variants, the control measures applied and the inclusion of external covariates. Therefore, the objective of this article is to contrast the efficiency of different models used in the literature for the analysis in predicting infection cases, considering only the data collected by the Ministry of Public Health in Mexico [16], through the use of time series.

3. Methodology used in the comparison of prediction models

The methodology was divided into five stages (Fig. 1). In the first, COVID-19 data from Mexico were collected; in the second, they were analyzed and pre-processed to represent them; in the third, they were organized by time series; in the fourth, various prediction models found in the literature were tested. The last stage was the graphing and comparison of the experimentation on the data generated daily during the pandemic.

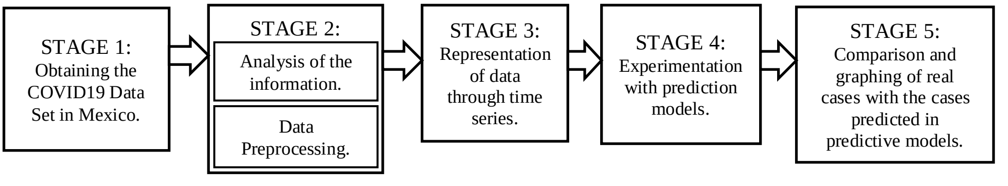

Fig. 1. Methodology for the analysis of time series with prediction models.

3.1 Obtaining the dataset

The data [16] comes from the Ministry of Health of the Government of Mexico, through the General Directorate of Epidemiology, and was reported during the period from April 19th, 2020 to December 31st, 2021. The data contains only the reported cases of COVID-19 in the different Health Centers nationwide distributed in the thirty-two states. The total number of records obtained was 12,698,740, of which 1,986,260 were analyzed, corresponding only to the states of Colima and the State of Mexico. In addition, each of the records can be divided into groups of: infected, deceased, municipality, state, latitude and longitude. Data were fitted to 7- and 30-day time series.

3.2 Analysis and preprocessing of information

Given the nature of the pandemic, the COVID-19 dataset [16] was found to have noise (mis captured information), redundancy, some missing values (mainly in the comorbidity fields) and outdated data, so it was proposed to normalize and transform in such a way that the resulting set was consistent. For the case study of this article, data corresponding to two states in Mexico were selected:

- *Colima:* the state with the smallest population (731,391 inhabitants) [17].
- *State of Mexico:* the state with the largest population (16,992,418 inhabitants) [17].

The state of Colima has a territorial area of 5,625 km^2 while the State of Mexico has 22,500 km^2, that is, 23.23 times larger than Colima.

During the year 2020, in the state of Colima 630,204 possible cases of contagion were reported (Table 1), but only 8,025 were confirmed (Table 2); while in the state of Mexico of the 5,832,576 cases reported (Table 1), only 161,809 infections were confirmed (Table 2), within the Epidemiological Surveillance System for Viral Respiratory Disease [16]. For the year 2021, the number of infection cases in Colima was 33,509 confirmed (Table 2), and for the State of Mexico there were only 427,068 (Table 2).

Table 1. Characteristics of the states of Colima and the State of Mexico [16, 18].

Description	Colima	State of Mexico
Number Inhabitants	731,391	16,992,418
Reported Cases Year 2020	630,204	5,832,576
Reported Cases Year 2021	658,977	11,126,420

Table 2. Data reported from the states of Colima and State of Mexico at the national level [16].

Types of Accumulated Cases	Colima	Estado De México
Confirmed 2020	8,025	161,809
Negatives 2020	7,438	210,710
Suspects 2020	1,802	56,530
Deaths 2020	820	23,961
Confirmed 2021	33,509	427,068
Negatived 2021	44,965	864,552
Suspects 2021	2,554	95,033
Deaths 2021	2,041	45,443

According to the number of cases registered at the national level (Table 1), Colima reported until 2021 almost 90% of cases with respect to its total population, and the State of Mexico up to 65%. A possible cause for the reporting of a greater number of cases in the state of Colima may be related to population density and its territorial distribution, since Colima has 130 inhabitants/km^2 on average, while the state of Mexico has 760. inhabitants/km^2, the latter concentrating the largest population in its capital with respect to its 125 municipalities.

Once the data was reviewed and normalized, preprocessing was done, for this a Univariate analysis was carried out in R and Python, to understand the distribution, central tendency, dispersion, and other aspects of each variable in the data set, through a statistical summary and application of Filter Methods such as ANOVA and Chi-Square [19]. In this stage, the objective was that each of the characteristics obtained in the data set were within the same scale (for example, in the sex attribute, the categorical values were 1 = woman, 2 = man and they jumped up to category 99 = not specified), with the purpose that the values of its distribution and frequency did not affect the interpretation of the predictive model. Some values were also imputed, such as the age attribute, where sometimes atypical data were found but since they were scarce, values were assigned using the median. Subsequently, SHAP (SHapley Additive exPlanations) [20] was used, because SHAP considers all possible combinations of features and calculates the average marginal contribution of each feature across these combinations, thus identifying which attribute is most important in a prediction and detecting biases or inconsistencies in the predictive model. For our case study, this tool showed that some attributes were irrelevant (such as those referring to comorbidities), helping to reduce the dimensionality of our dataset. Finally, we verified the homoscedasticity of the data with the Box-Cox Transformation. For future experiments, the accumulated cases were counted to obtain the number of confirmed cases of infection, negative cases, suspected cases, deaths, and recovered cases.

3.3 Representation of data using time series

A crucial tool to understand, manage and control the spread of a disease is the time series [21]; to represent the information from the states of Colima and Mexico, the total number of infection cases reported during the years 2020 and 2021 was counted by daily date and by week (Fig. 2 and 3).

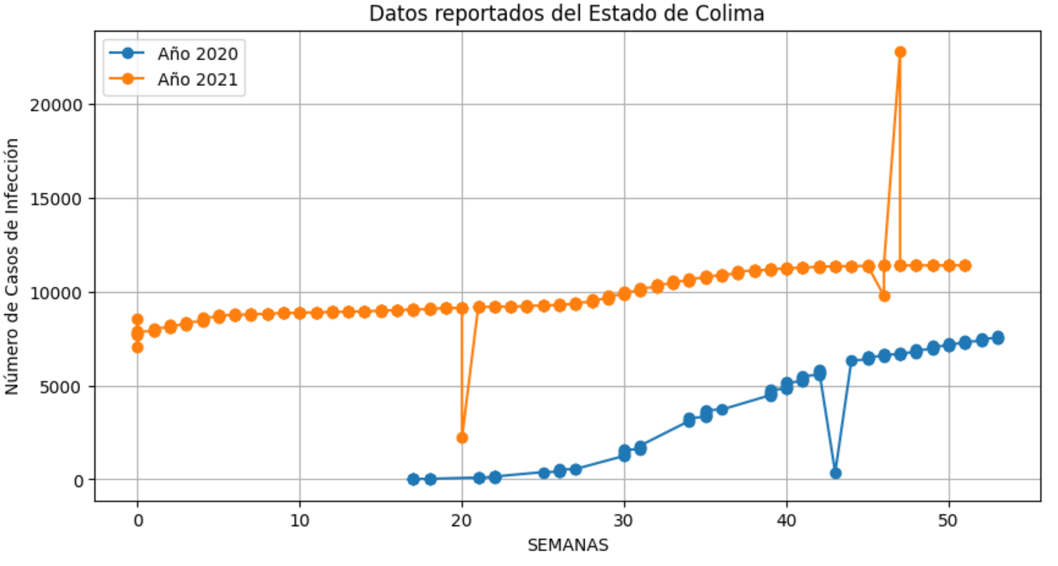

Fig. 2. Time series of accumulated cases of COVID-19 infection in the state of Colima, period April 2020 to December 2021.

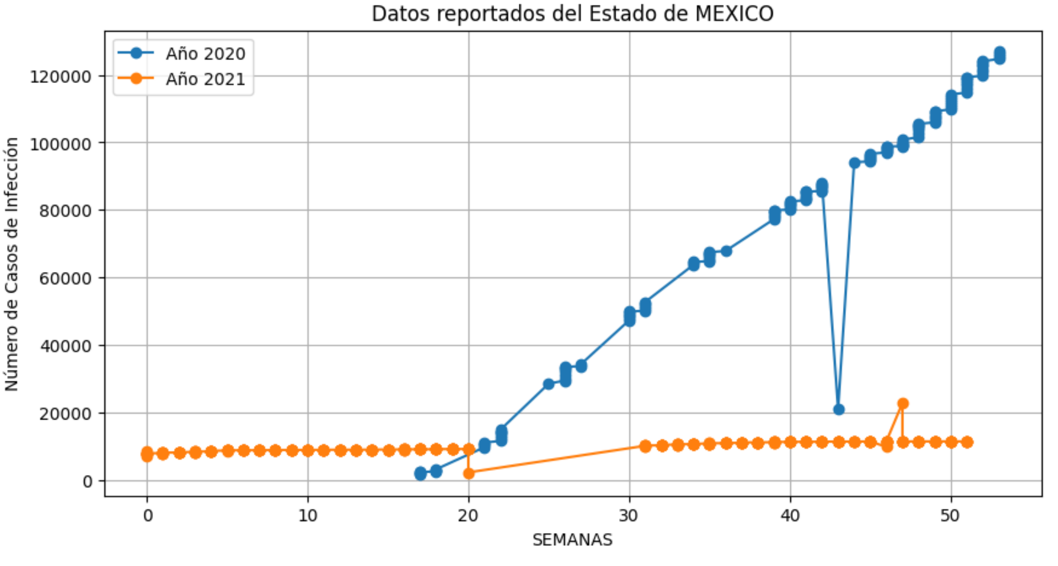

Fig. 3. Time series of accumulated cases of COVID-19 infection in the State of Mexico, period April 2020 to December 2021.

In the two previous series, the similarity can be seen in week 42 of the year 2020 (January 4th, 2021 and January 6th, 2021) and week 20 of the year 2021 (January 4th, 2021 and January 7th, 2021) where a downward peak is observed, this is explained because changes were made to the records, and new attributes were added: PCR test information and new classifications. In week 45 of 2021 year, Colima registered a decrease due to the cases of deaths reported compared to the cases of infection;

In week 46 of 2021 year (corresponding to the dates from November 15th, 2021 to November 21st, 2021), both states had an uptick in contagion cases. The above could have derived from the case of omission in the containment measures with respect to the previous two weeks, since in Mexico the month of November is a month of traditions, in addition to the fact that approximately 80% of the population was already vaccinated with the first dose [22]. In the distribution of the data for both states, a certain seasonal component can be understood, because a stable range without trend is maintained in the periods of growth and decrease. Given the complexity of the problem addressed, in terms of the amount of information collected, the trends were exemplified on an annual, monthly, and weekly basis, taking periods of 7 days and 30 days for experimentation.

4. Results in the comparison of prediction models

4.1 Experimentation with prediction models

In the literature, various models have been found that address the problem of estimating and/or predicting the number of infection cases through time series [23] [24]. Therefore, six models were experimented with that used different criteria, concepts, and methodologies, classifying them as deterministic and stochastic (Fig. 4). In the deterministic model [25] the values in its parameters are usually constant and rigid (controlled behavior), while in the stochastic model [25] the parameters used have random or estimated values (probabilistic behavior), that is, according to the literature, the latter captures the behavior of diseases more realistically. Once the data and predictive models were identified, they were compared and experimented with each of them.

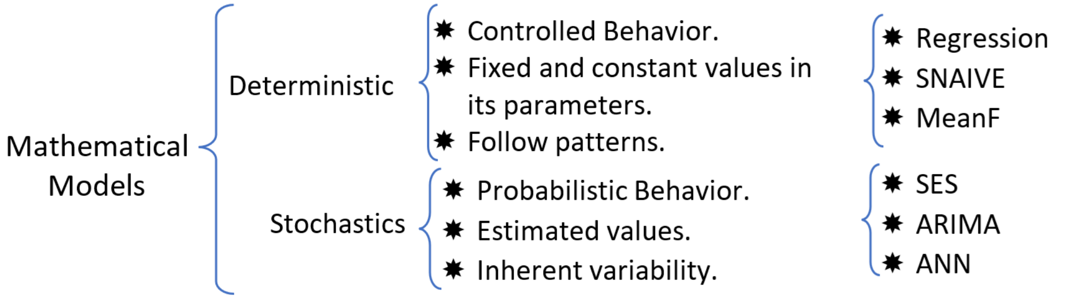

Fig. 4. Classification of mathematical models in the prediction of COVID-19 infection cases.

4.2 Comparison between Regression, SNAIVE and MeanF

The regression model [26-27] allows the incorporation of various predictor variables, however, it attempts to place the value of its parameters in a linear manner, which makes its implementation difficult to represent the real world, that is, collinearity in it can lead to imprecise estimates, so their confidence intervals are wide. The SNaive (Simple Smoothing) [28] and MeanF (Averaging Method) [29-30] models are methods based on past values of a time series, so their prediction focuses on sequential data, so they do not adequately capture the trend. seasonality and randomness. noise. These models are represented with the following equations (Table 3).

As seen in Fig. 5, using the Regression model [27], the prediction does not converge with almost any data, however, it seems to exemplify the trend of the data. In Fig. 6, for the State of Mexico, the regression model simply did not reflect a prediction during the year 2020 with an observed period of 7 days; and for the data for the year 2021, it only reflected the trend they had.

Since the SNaive model [28] focuses on the estimated value of the average of the previous values, in the graphs of both the state of Colima (Fig. 7) and the state of Mexico (Fig. 8) similarities were observed in the behavior of the real values with the predicted ones, although the latter are far from the real data.

Table 3. Equations of Regression, SNaive y MeanF models.

Model	Equation		Parameters
Regression	$Y = \beta_0 + \beta_1 X + \varepsilon$	(5)	Y is dependent variable (that wants it to predict). β_0 is the ordinate at the origin or intercept. β_1 is the slope of a regression line. X is the independent variable (used to make the prediction). ε is the error that represents the variability not explained by the model.
SNAIVE	$\check{Y}_{t+1} = Y_t$	(6)	$\check{Y}$ represents the prediction for time $t+1$. Y_t is the last value observed in time. t time.
MeanF	$\check{Y}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i$	(7)	$\check{Y}_{t+1}$ represents the prediction for time $t+1$. Y_i are values observed at the previous times, from $i = 1$ to $i = t$. t is the number of past observations in the time series.

The MeanF model [30], as its name indicates, is based on calculating the arithmetic mean of past values in a time series to predict the future value, so it follows a constant pattern based on historical observations, that is, its relationship with time series lies in making forecasts on sequential data. For the infection cases from the state of Colima (Fig. 9), given that the data was scarce, only when these were described every 7 days was it possible to capture a pattern similar to that of the real data, however, in obtaining predictive data at 30 days, this model did not even exemplify a trend. With the experimentation of the MeanF model in the state of Mexico (Fig. 10), during 2020 year the data observed for 7 days had wider fluctuations compared to the real data; for the information collected in 2021, MeanF became more robust and less sensitive to changes, reducing errors between observed and predicted data.

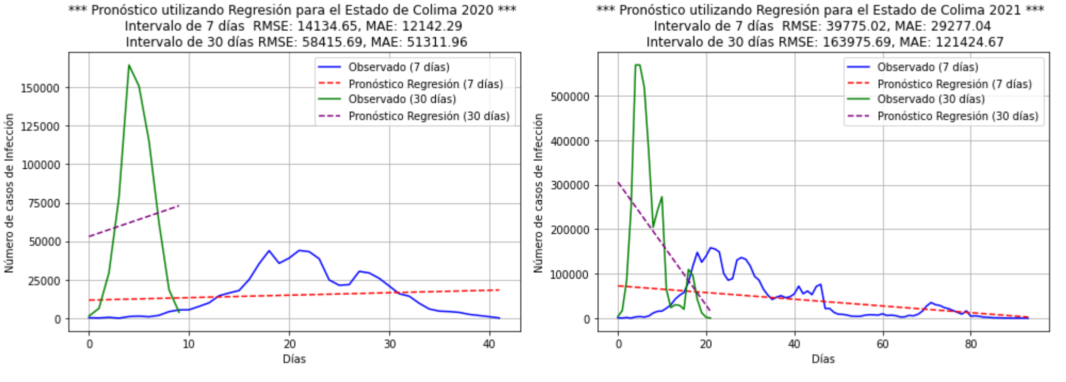

Fig. 5. Regression Model of the state of Colima.

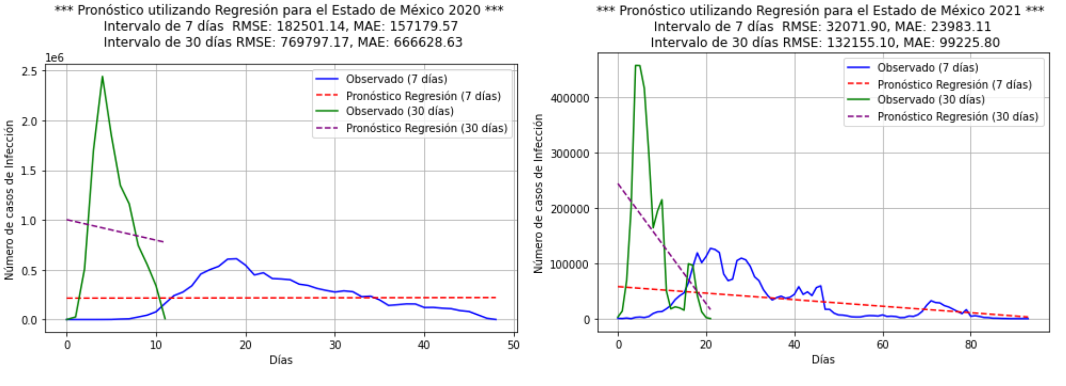

Fig. 6. Regression Model of the state of Mexico.

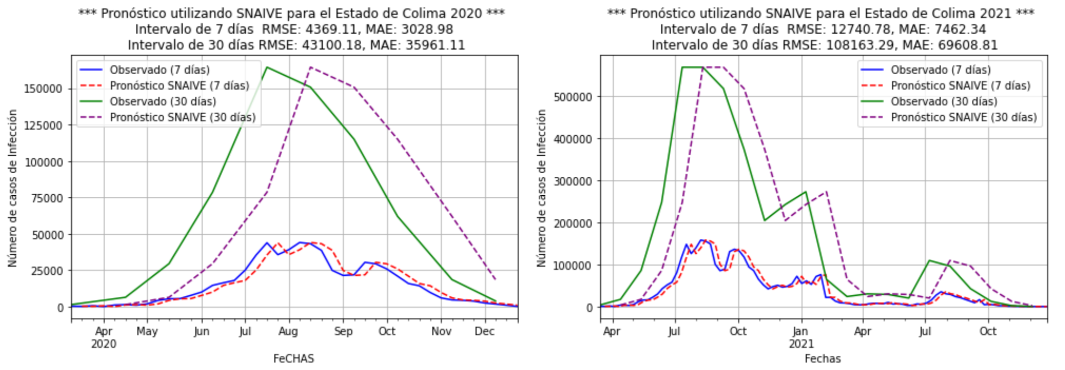

Fig. 7. SNAIVE model of the state of Colima.

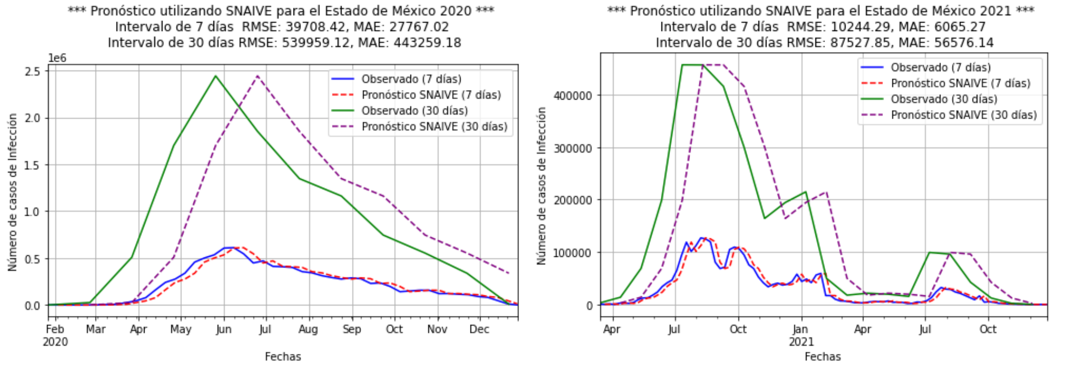

Fig. 8. SNAIVE model of the state of México.

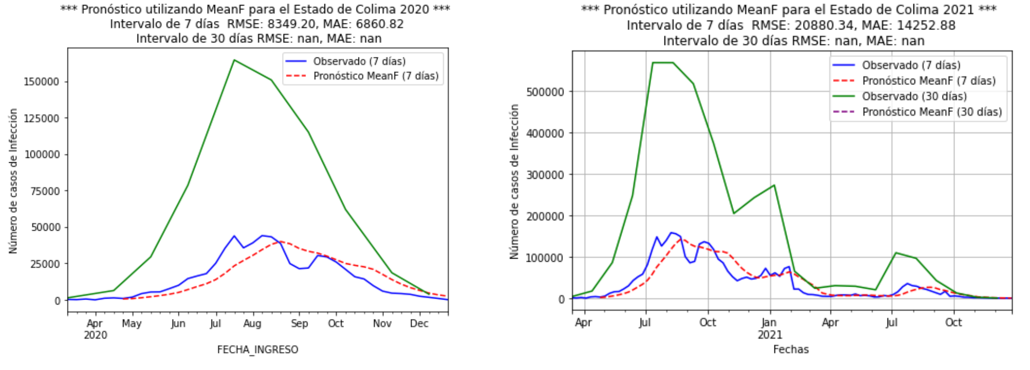

Fig. 9. MeanF model of the state of Colima.

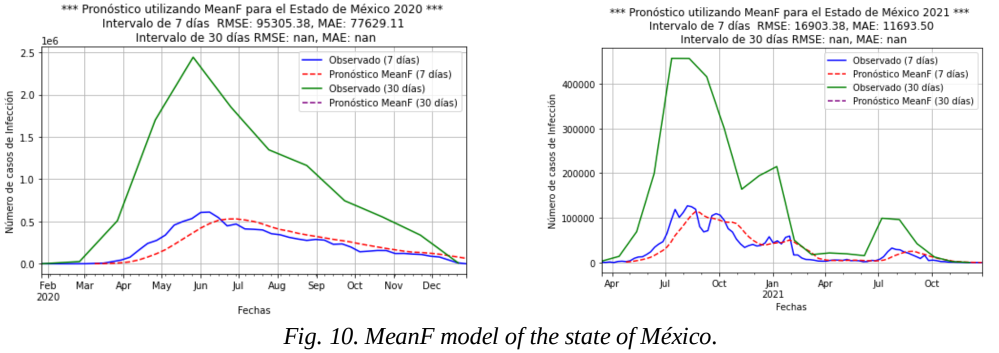

Fig. 10. MeanF model of the state of México.

4.3 Comparison between SES, ARIMA, ANN

The SES [29], ARIMA [30] and ANN [31-32] models are of the stochastic type and explicitly capture the variability of past information, that is, they estimate their predictions from a tuning of the errors generated from historical data, which generates greater certainty. These models are represented by the equations shown in Table 4.

Table 4. Equations of the SES, ARIMA y ANN models.

Model	Equation	Parameters
SES	$\check{Y}_{t+1} = \alpha \cdot Y_t + (1-\alpha) \cdot \check{Y}_t$ (8)	$\check{Y}_{t+1}$ is the prediction for time $t+1$. Y_t is the value observed at time t. $\check{Y}_t$ is the prediction for time t (the predicted value in the previous period). α is the smoothing factor, a value between 0 and 1 that controls the weight of the most recent observation to the previous prediction.
ARIMA	$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$ (9)	Y_t is the value of the time series at time t. c is a constant $\phi_1, \phi_2, \dots, \phi_p$ are the autoregression parameters, which represent the relationship between the current value and the past values of the series. $\theta_1, \theta_2, \dots, \theta_q$ are the moving average parameters, which represent the relationship between the current value and the past errors of the series. ε_t is the error term at time t, reflecting variability not explained by the autoregressive and moving average components of the model.
ANN	$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$ $a^{[l]} = g(z^{[l]})$ (10)	$z^{[l]}$ is the activation vector before applying the activation function at layer l. $W^{[l]}$ is the weight matrix associated with layer l. $a^{[l-1]}$ is the activation vector of the previous layer. $b^{[l]}$ is the bias vector at layer l. $g()$ is the activation function applied element by element.

Although the SES [29] model is easy to implement compared to others such as ANN [31, 33] and ARIMA [30, 32], in this experimentation, one of its limitations in prediction was the need to assign an alpha value, which controls the most recent observation, which leads to a lack of long-term memory, consequently it falls into an analysis where the trend of the data must be more linear or constant, no matter if the amount of information is little (Fig. 11) or a lot (Fig. 12) to predict. In this experimentation, the SES model simply did not converge with any of the data.

The ARIMA (Autoregressive Integrated Moving Average) model [31] consists of 3 main components:

- AR (Autoregressive): which models the dependence of a current observation on past observations.
- I (Integrated): which indicates the stationarity of a time series.
- MA (Moving Average): indicates the relationship of the dependence between a current observation and past errors, using moving average coefficients.

In the ARIMA experiment, the time series of the state of Colima (Fig. 13) and the State of Mexico (Fig. 14) presented the lowest error, likewise, the prediction reflected a more adequate behavior concerning the real data.

Finally, an ANN MLP (Multi-Layer Perceptron) model [32] was experimented with, which, by its nature, can solve problems that are not linearly separable in addition to capturing complex relationships between variables. In this experiment, for the case of the state of Colima (Fig. 15), the predicted data matched the real data better when they were studied for 7 days.

With the data from the state of Mexico (Fig. 16), a better result will also be considered when these are analyzed in a shorter time.

Fig. 11. SES model of the state of Colima.

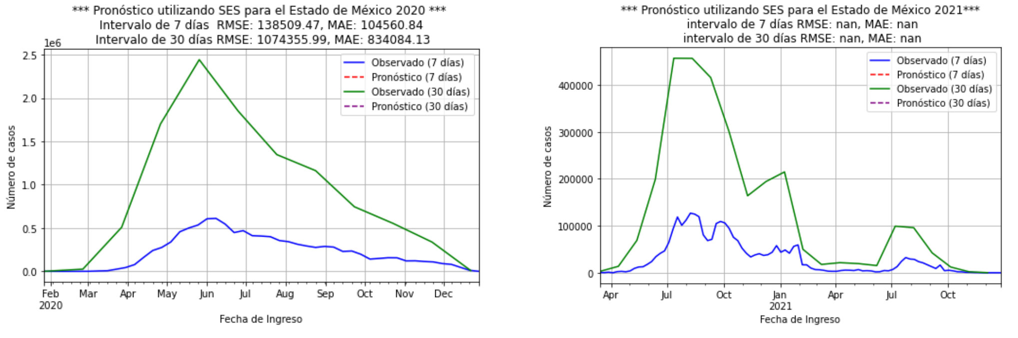

Fig. 12. SES model of the state of México.

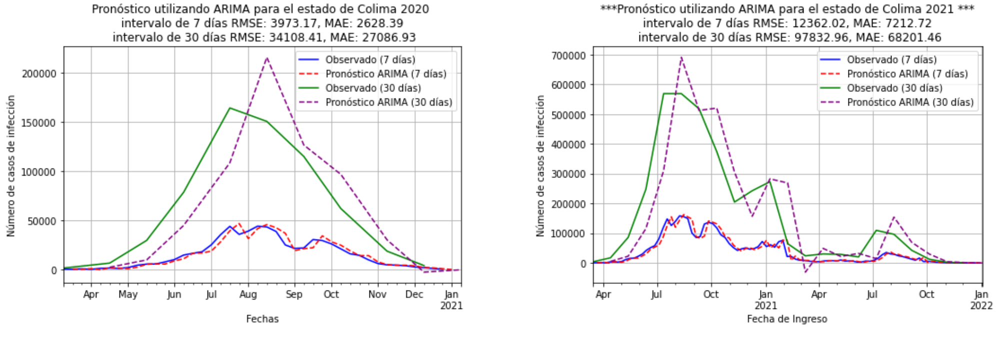

Fig. 13. ARIMA model of the state of Colima.

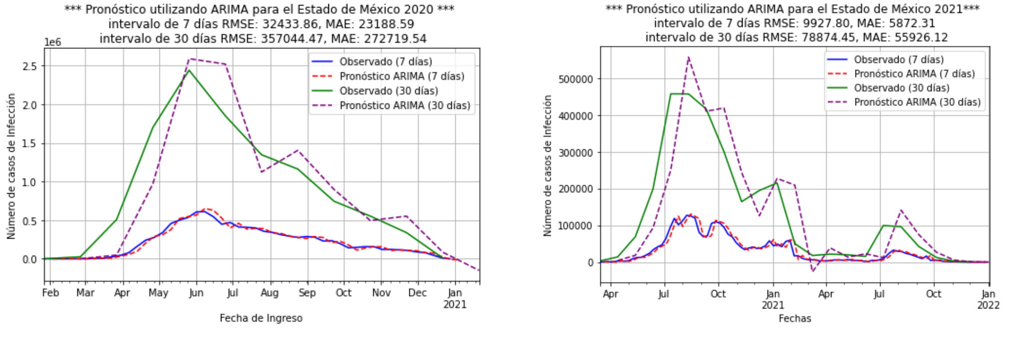

Fig. 14. ARIMA model of the state of México.

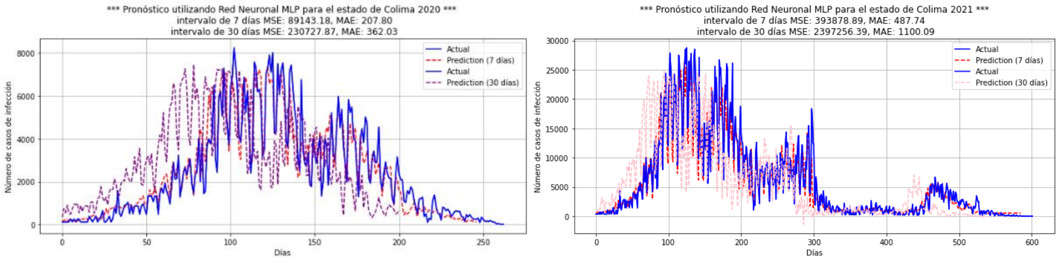

Fig. 15. MLP model of the state of Colima.

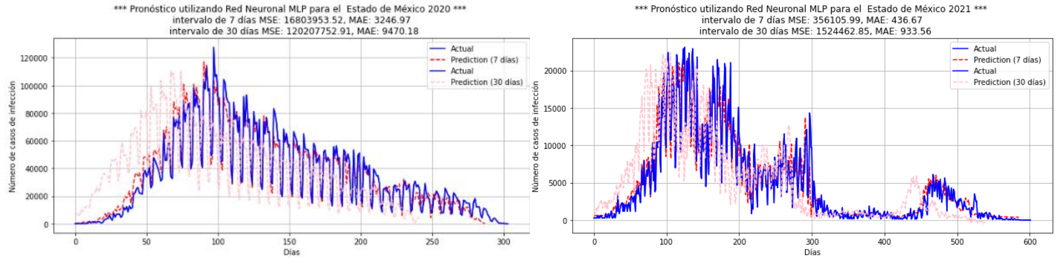

Fig. 16. MLP model of the state of México.

4.4 Discussion of results

To evaluate each of the predictive models, the averaged error biases were obtained (from the real data with the predicted data), and they were evaluated with the RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) metrics. A low RMSE indicates that the model is producing accurate predictions, however, because errors are first squared before averaging, this metric comes to penalize large errors more. Likewise, if an MAE is low, it indicates that the model has accurate predictions, which, due to its nature, only shows the difference between the predicted value and the actual value. The following Tables (Tables 5 and 6) show the data resulting from the application of these metrics.

Table 5. Description of the predictive models in the time series of the State of Colima.

State | Year	COLIMA 2020				COLIMA 2021			
Period	7 days		30 days		7 days		30 days	
Model | Error	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Regression	14,134.65	*12,142.29*	58,415.69	*51,311.96*	39,775.02	*29,277.04*	163,975.69	*121,424.67*
SNAIVE	4,369.11	3,028.98	43,100.18	35,961.11	12,740.78	7,462.34	108,163.29	69,608.81
MeanF	8,349.20	6,860.82	NAN	NAN	20,880.34	14,252.88	NAN	NAN
SES	NAN	NAN	NAN	NAN	NAN	NAN	NAN	NAN
ARIMA	3,973.17	2,628.39	34,108.41	27,086.93	12,362.02	7,212.72	97,832.96	68,201.46
MLP	*89,143.18*	**207.80**	*230,727.87*	**362.03**	*393,878.89*	**487.74**	*2,397,256.39*	**1,100.09**

**Notation: Numbers in italics indicate high error values. Shaded values indicate low values. Values in bold indicate the lowest values. NAN notes that the model did not make the prediction.*

Table 6. Description of the predictive models in the time series of the State of Mexico.

State | Year	STATE OF MÉXICO 2020				STATE OF MÉXICO 2021			
Period	7 days		30 days		7 days		30 days	
Model | Error	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Regression	182,501.14	*157,179.57*	769,797.17	666,628.63	32,071.90	*23,983.11*	132,155.10	*99,225.80*
SNAIVE	39,708.42	27,767.02	539,959.12	443,259.18	10,244.29	6,065.27	87,527.85	56,576.14
MeanF	95,305.38	77,629.11	NAN	NAN	16,903.38	11,693.50	NAN	NAN
SES	138,509.47	104,560.84	1,074,355.99	*834,084.13*	NAN	NAN	NAN	NAN
ARIMA	32,433.86	23,188.59	357,044.47	272,719.54	9,927.80	5,872.31	78,874.45	55,926.12
MLP	*16,803,953.52*	**3,246.97**	*120207752.91*	9,470.18	*356,105.99*	**436.67**	*1,524,462.85*	**933.56**

**Notation: Numbers in italics indicate high error values. Shaded values indicate low values. Values in bold indicate the lowest values. NAN notes that the model did not make the prediction.*

Regarding the period analyzed, the lowest errors occurred in the data analyzed for 7 days. The individual results of the MeanF and SES models showed that they were not able to predict data at some point in the experimentation, because they require controlling or averaging previous data so that the prediction has a value between 0 and 1. The results of the Regression model frequently presented the highest error values in the MAE metric, which indicates that the predicted data never converged with the real data.

The SNAIVE model, although it presented errors, the predicted data in its graph showed behavior similar to the real data. The values with the lowest errors, shown in Tables 5 and 6, concerning the experimental models were those of ARIMA and MLP (shaded in gray); Although the ANN MLP came to present the highest values of errors in the RMSE metric, this only tells us that it usually adjusts/adapts better to a longer period of time in addition to reducing the error more significantly at a greater number. of data. The lowest results were presented by the MLP model (shaded in Bold), with an observed period of 7 days. Within the previous graphs (section 3), you can see more clearly the projection of the real data and the predicted data.

5. Conclusions and future work

Taking into consideration the set of data provided by the Ministry of Health of Mexico, some of the predictive models did not show the expected predictive behavior; however, the results shown by ARIMA and ANN showed lower errors compared to the other models. The prediction analysis carried out in this work contrasts with the results of other investigations; unfortunately, most of these do not share the parameters used, and the results usually vary concerning the data set they study in their experimentation, which makes it difficult to improve the models and/or be used as a reference to continue with new research. With the results obtained in this research, it is proposed to develop a model that focuses on prediction, not only of cases of infection of the COVID-19 pandemic but also of other diseases, adding other factors and/or experimenting with other types of models of ANN like the LSTM.

Code Availability

All results for this study were generated using R and Python. The code that generated the figures in this article is available at https://github.com/KeilaVCortes/COVID19_prediction.

References

[1]. Organización Mundial de la Salud (OMS), “Enfermedad por el Coronavirus,” Novel Coronavirus 2019, 42–49 (2021). [Last access: February 28th, 2021]. https://www.paho.org/es/enfermedad-por-coronavirus-covid-19.

[2]. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), “Línea de tiempo COVID-19 en México”, [Last access: December 27th,2023]. https://micrositios.inai.org.mx/conferenciascovid-19tp/?page_id=8432.

[3]. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), “Sugerencias de Actividades para realizar durante la contingencia sanitaria #Quédate en Casa”, pp. 1-45 (2021). https://inicio.inai.org.mx/nuevo/RecomendacionesINAICOVID-19.pdf.

[4]. J. C. Clement, V. K. Ponnusamy, K. Sriharipriya, et al., “A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis”, IEEE reviews in biomedical engineering, 15, 325–340(2022). https://doi.org/10.1109/RBME.2021.3069213.

[5]. A. Mehrotraa y R. Agarwal, “A Review of Use of Data Mining duringCOVID-19 Pandemic”, Turkish Journal of Computer and Mathematics Education, 12(6), 4547-4552(2021). https://doi.org/10.17762/turcomat.v12i6.8441.

[6]. Vasilev, I.A., Petrovskiy, M.I., Mashechkin, I.V, et al. "Predicting COVID-19-Induced Lung Damage Based on Machine Learning Methods". Programming and Computer Software, 48, 243–255 (2022).doi: https://doi.org/10.1134/S0361768822040065.
[7]. F. Rustam, A., et al. "COVID-19 Future Forecasting Using Supervised Machine Learning Models", IEEE Access Xplore, 8, 101489-101499(2020). doi:10.1109/ACCESS.2020.2997311.
[8]. Madini O. Alassafi, Mutasem Jarrah, Reem Alotaibi, "Time series predicting of COVID-19 based on deep learning", Neurocomputing 468, 335-344 (2022). ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.10.035.
[9]. R. Chandra, A. Jain y D. Singh Chauhan, "Deep learning via LSTM models for COVID-19 infection forecasting in India", PLOS ONE, 17(1), 1-28 (2022), https://doi.org/10.1371/journal.pone.0262708.
[10]. S. Kumar, R. Sharma, T. Tsunoda, et al., "Forecasting the spread of COVID-19 using LSTM network". BMC Bioinformatics, 22 (6), 316 (2021). https://doi.org/10.1186/s12859-021-04224-2.
[11]. Z. Car, S. Baressi Šegota, N. Anđelić, et al., "Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron", Hindawi. Computational and Mathematical Methods in Medicine, 2020 (5714714), 1-10 (2020). https://doi.org/10.1155/2020/5714714.
[12]. R. Somyanonthanakul, K. Warin, W. Amasiri, et al., "Forecasting COVID-19 cases using time series modeling and association rule mining", BMC Medical Research Methodology, 22(281), 1-18 (2022), doi: 10.1186/s12874-022-01755-x.
[13]. M. Abdelaziz, A. Ahmed, A. Riad, et al., "Forecasting daily confirmed COVID-19 cases in Algeria using ARIMA models", Short research communication- EMHJ, 29 (7), 515-519 (2023), https://doi.org/10.26719/emhj.23.054.
[14]. S. Sah, B. Surendiran, R. Dhanalakshmi, et al., "Forecasting COVID-19 Pandemic Using Prophet, ARIMA, and Hybrid Stacked LSTM-GRU Models in India", Hindawi. Computational and Mathematical Methods in Medicine, 2022 (1556025), 1-19 (2022), https://doi.org/10.1155/2022/1556025.
[15]. Y. Jin, R. Wang, X. Zhuang, et al., "Prediction of COVID-19 Data Using an ARIMA-LSTM Hybrid Forecast Model", MDPI Mathematics, 10 (21), 1-13 (2022), https://doi.org/10.3390/math10214001.
[16]. Dirección General de Epidemiología, "Datos Abiertos Bases Históricas". Available at.: https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia, [Last access: January 25th, 2024].
[17]. Instituto Nacional de Estadística y Geografía (INEGI), "Consulta de indicadores sociodemográficos y económicos por área geográfica". Available at.: https://www.inegi.org.mx/, [Last access: Decembre 27th,2023].
[18]. Instituto Nacional de Estadística y Geografía (INEGI) para niños. "Densidad de Población", Available at.: https://cuentame.inegi.org.mx/poblacion/densidad.aspx?tema=P, [Last access: Decembre 27th,2023].
[19]. G. Ramírez-Valverde y B. Ramírez-Valverde, "Modelo estadístico para defunciones y casos positivos de COVID-19 en México", EconoQuantum, 18(1), 1-20 (2021). https://doi.org/10.18381/eq.v18i1.7223.
[20]. Rojas-García, M., Vázquez, B., Torres-Poveda, K.et al. "Lethality risk markers by sex and age-group for COVID-19 in Mexico: a cross-sectional study based on machine learning approach". BMC Infectious Diseases, 23(18), 1-14 (2023). https://doi.org/10.1186/s12879-022-07951-w.
[21]. Cueva Moncayo, M., Jami Carrera, J., & Betancourt Rubio, E., "Previsión de COVID-19 mediante modelo de series temporales en pacientes de una clínica de salud rural", Revista Cubana de Investigaciones Biomédicas, 42(3127), 1-21 (2023), Editorial de Ciencias Médicas (ECIMED). https://revibiomedica.sld.cu/index.php/ibi/article/view/3127.
[22]. Secretaría del Gobierno de México (gob.mx), "Boletín estadístico 2021 del exceso de mortalidad por todas las causas durante la emergencia por COVID-19", Grupo Interinstitucional para la estimación del exceso de mortalidad por todas las causas, 46 (27), 1-42 (2021). https://coronavirus.gob.mx/wp-content/uploads/2021/09/2021.09.08-Boleti%CC%81n_XXI_Exceso_Mortalidad_SE33_MX.pdf.
[23]. Z. Malki, E.-S. Atlam, A. Ewis, et al., "The COVID-19 pandemic: prediction study based on machine learning models", Environmental Science and Pollution, 28(30), 40496-40506 (2021). doi: 10.1007/s11356-021-13824-7.
[24]. L. Tomov, L. Chervenkov, D. G. Miteva, et al. "Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic", WJCC (World Journal of Clinical Cases), 11 (29), 6974-6983 (2023). PMID: 37946767 doi: 10.12998/wjcc.v11.i29.6974.
[25]. M. S. Bartlett, "Deterministic and Stochastic Models for Recurrent Epidemics", Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, IV, 81-109 (1956). https://apps.dtic.mil/sti/pdfs/AD1028592.pdf.

[26]. A. M. Radwan, “Forecasting of Covid-19 Using Time Series Regression Models”, 2021 Palestinian International Conference on Information and Communication Technology (PICICT), 7-12 (2021). doi: 10.1109/PICICT53635.2021.00014.

[27]. S. Bhutia, B. Patra y M. Ray, “COVID-19 epidemic: analysis and prediction”, IAES International Journal of Artificial Intelligence (IJ-AI), 11(2), 736-745 (2022). ISSN: 2252-8938, doi: 10.11591/ijai.v11.i2.

[28]. V. Chaurasia y S. Pal, “Application of machine learning time series analysis for prediction COVID-19 pandemic”, Research on Biomedical Engineering. Sociedade Brasileira de Engenharia Biomedica, 38(1), 35-47 (2020) . PMCID: PMC7585491. doi: 10.1007/s42600-020-00105-4.

[29]. Fadhil A. Abidi; Zainb Hasan Radiy, “Using exponential smoothing methods to analysis COVID-19 time series”, 2nd International Conference on Engineering and Science to Achieve the Sustainable Development Goals. AIP Conference Proceedings, 3092 (1), id.070014, 1-10 (2024), https://doi.org/10.1063/5.0200417.

[30]. G. Sanglier Contreras, M. Robas Mora, P. Jimenez Gomez, "Use of quantitative forecasting methods and error calculation for better adaptability to the application of a mathematical model to determine the speed of spread of a coronavirus infection (COVID-19) in Spain", Contemporary Engineering Sciences, 13(1), 157-175 (2020), doi: 10.12988/ces.2020.91570.

[31]. R. Somyanonthanakul, K. Warin, W. Amasiri, et al. “Forecasting COVID-19 cases using time series modeling and association rule mining”, BMC Medical Research Methodoly, 22 (281), 1-18 (2022), https://doi.org/10.1186/s12874-022-01755-x.

[32]. Yu, C. S., Chang, S. S., Chang, et al., “A COVID-19 Pandemic Artificial Intelligence-Based System with Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study”, Journal of medical Internet research, 23(5), (2021). e27806. https://doi.org/10.2196/27806.

[33]. Kadirliev T.M., Telpukhov D.V., Solovyev R.A. Application of Neural Networks for Image Segmentation in the Problem of Fast Global Routing. Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). 2023;35(5):145-156. (In Russ.) https://doi.org/10.15514/ISPRAS-2022-35(5)- 10.

Информация об авторах / Information about authors

Кейла Басти КОРТЕС-МАРТИНЕС в 2019 году получила степень магистра компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика). Ее научные интересы – машинное обучение, распознавание образов и интеллектуальный анализ данных.

Keila Vasthi CORTÉS-MARTÍNEZ received a master’s degree in computer science from the National Center for Research and Technology Development of Cuernavaca, México, in 2019. Her research interests are machine learning, image recognition and data mining.

Хуго ЭСТРАДА-ЭСКИВЕЛ получил степень магистра в области компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика) и степень PhD в области информатики в Техническом университете Валенсии (Испания), а также степень PhD в области информатики и телекоммуникаций в Университете Тренто (Италия) в 2008 году. Научный сотрудник исследовательского центра INFOTEC, Национального совета по науке и технологиям и в настоящее время в Национальном центре исследований и технологических разработок CENIDET. Его научные интересы включают большие данные, интернет вещей и умные города.

Hugo ESTRADA-ESQUIVEL received a master’s degree in Computer Science from the National Center for Research and Technology Development of Cuernavaca, México, and a PhD degree in Informatics from the Technical University of Valencia, Spain, and a Ph.D. degree in Informatics and Telecommunications from the Trento University Italy, in 2008. PhD Estrada has been researcher in INFOTEC research center, the National Council of Science and Technology and currently in the National Center for Research and Technological Development, CENIDET. His research interests include big data, Internet of things and smart cities.

Алисия МАРТИНЕС-РЕБОЯР получила степень магистра в области компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика) и степень PhD в области информатики в Техническом университете Валенсии (Испания), а также степень доктора философии в области информатики и телекоммуникаций в университете Тренто (Италия) в 2008 году. С 2009 года она является профессором-исследователем Национального центра исследований и технологических разработок CENIDET. Автор 4 книг и более 130 работ. Ее научные интересы включают большие данные, Интернет вещей, умные города и аффективные вычисления.

Alicia MARTINEZ-REBOLLAR received her master degree in Computer Science from the National Center for Research and Technology Development of Cuernavaca, México, and a PhD degree in Informatics from the Technical University of Valencia, Spain, and a PhD degree in Informatics and Telecommunications from the Trento University Italy, in 2008. Since 2009, she has been Research professor with the National Center for Research and Technological Development, CENIDET. She is the author of 4 books, and more than 130 papers. Her research interests include big data, Internet of things, smart cities, and affective computing.