Proc. of the 8th International Conference on Engineering and Emerging Technologies (ICEET) 27-28 October 2022, Kuala Lumpur, Malaysia

2022 International Conference on Engineering and Emerging Technologies (ICEET) | 978-1-6654-9106-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEET56468.2022.10006845

Predicting Covid-19 Cases for 12 Countries using Long Short-Term Memory

Pranav Ramesh
Department of Computer Science,
Birla Institute of Technology and
Science Pilani, Dubai Campus,
Dubai, United Arab Emirates.
f20180089@dubai.bits-pilani.ac.in

J. Angel Arul Jothi
Department of Computer Science,
Birla Institute of Technology and
Science Pilani, Dubai Campus,
Dubai, United Arab Emirates.
angeljothi@dubai.bits-pilani.ac.in

***Abstract*— A novel virus named coronavirus or 'COVID-19' by the World Health Organization (WHO) has spread around the entire world placing mankind in a situation that no one had predicted. The rise of the number of infected and death cases around the world is alarming and has caused hysteria among mankind. Considering the adversity of the COVID-19, some immediate plan to monitor the number of cases in the future needs to be maneuvered. In this paper, we aim to implement a method to envision the number of COVID-19 cases for the future. We achieve the result by using a deep learning algorithm known as Long Short-Term Memory (LSTM) over the real-time dataset provided by WHO for predicting the number of COVID-19 cases in twelve countries. The countries considered in this study are United States of America, China, United Arab Emirates, India, Brazil, France, Germany, Spain, Republic of Korea, Italy, Singapore, and Argentina. The contribution of this paper is to provide each country with their own model that can help predict their respective future COVID-19 cases. With these predictions, each country can then come up with solutions to reduce the number of infected cases in their respective nation. The proposed LSTM model was evaluated using metrics such as Correlation Coefficient and R^2 Error. The results show that the model was giving high R^2 score (≥ 0.7) and high correlation coefficient (≥ 0.7) between the test and train datasets. In the cases where R^2 score (< 0.7) and correlation coefficient (< 0.7) were low, the train and test values of the datasets were similar making the predictions accurate.**

Keywords*— *Covid-19, Deep Learning, LSTM, Prediction, Forecasting

I. Introduction

The coronavirus disease is transferred by encountering an infected person and touching an infected substance or object. During the early stages of the propagation of this disease the WHO labelled it a pandemic, meaning that it is an extremely serious and deadly disease. The sudden and rapid spread of coronavirus has created considerable challenges, global imbalances, and issues to all communities. Manufacturers and hospitals across the world were not able to keep up with the demand for ventilators, personal protective equipment (PPE kits), and intensive care unit beds. Some healthcare providers started making ventilators and ICU units in their garages and parking spaces to deal with the massive surge of infections.

Researchers and doctors have failed to properly comprehend aspects associated to COVID-19, such as its transmission rate. Because of the coronavirus's unfamiliarity and novelty, contradictory messages were given about the outbreak's potential severity, the importance of wearing a mask, being socially distanced and other related precautions. This resulted in ad hoc and an increase in COVID-19 cases. These issues necessitated swift answers and solutions.

It is well known that the coronavirus has a considerable impact on people's health and can even cause death, either directly or by increasing pre-existing health conditions. Because the COVID-19 pandemic has afflicted a vast number of individuals around the world and there is no cure, it is critical to estimate the number of possible cases using current data. Due to the virus's novel origin and ability to create new strains, the process to accurately model and predict the infection rate is extremely difficult.

Many academics, especially data scientists, have been striving to fight COVID-19. Data scientists can contribute to the research by creating prediction models that emphasize the likely actions of the virus which can then be used to properly anticipate the virus's spread. Recently, deep learning (DL) models are thought to be tools that can aid in the development of prediction models. Even though several neural networks (NNs) have been reported in the past, the long short-term memory (LSTM) has been examined in recent times for forecasting or prediction purposes because it can employ temporal data.

Hence, in this research work, we use the LSTM model, a deep NN that learns long-term dependencies in data by memorizing information for long periods of time to predict the number of Covid-19 positive cases over time for twelve countries. For this purpose, the following countries are considered: United States of America, China, United Arab Emirates, India, Brazil, France, Germany, Spain, Republic of Korea, Italy, Singapore, and Argentina. We used the LSTM model because it was well-suited to the task.

TABLE I. LITERATURE REVIEW

Reference No	Aim and Objective	Methodology	Evaluation Metric	Results
[1]	To develop models that can be applied for real-time prediction of COVID-19 activity in all individual countries and territories worldwide.	Random forest regression algorithm, LSTM, Decision tree	Mean absolute error (MAE), Root mean square error (RMSE), Pearson correlation coefficient, and Spearman correlation coefficient	RMSE: 14.29 Pearson Correlation Coefficient: 0.76
[2]	Modelling cases of both confirmed diagnoses and deaths from COVID-19 in Chile using Autoregressive Integrated Moving Average models, Exponential Smoothing techniques, and Poisson models	Autoregressive Integrated Moving Average	Mean Error, RMSE, MAE, Mean percentage Error, Mean Absolute Percentage Error (MAPE)	RMSE: 1904.86
[3]	Short-term prediction of Covid-19 using past data.	Poisson and Gamma distribution modelling for prior probability and finding the posterior distribution for predictions.	From graphs	
[4]	Studies the transmission process of the Corona Virus Disease 2019 (COVID-19).	Monte Carlo	Graphs	
[5]	Predicting cumulative cases and deaths caused by Covid 19 for India, Italy, Japan, Spain, UK, and US	Auto-Regressive Integrated Moving Average and Long Short-Term Memory	Akaike Information Criteria, MAPE and RMSE	Max RMSE: 16,408.89
[6]	Indian states with COVID-19 hotpots and capture the first (2020) and second (2021) wave of infections and provide two months ahead forecast.	Long short-term memory (LSTM), bidirectional LSTM, and encoder-decoder LSTM	RMSE	Max RMSE ~ 35,000
[7]	In comparisons with real-time recurrent learning, back propagation through time, recurrent cascade correlation, Elman nets, and neural sequence chunking, LSTM leads to many more successful runs, and learns much faster.	LSTM	Graphs	
[8]	Method to identify whether a patient has risk of COVID-19 using Logistic Regression model, considering multiple symptoms.	Logistic Regression	Confusion Matrix	Accuracy: 0.92111
[9]	The proposed model was evaluated and compared with 17 baseline models on test and forecast data. The primary finding of this research is that the proposed CNN-LSTM model outperformed them all	CNN-LSTM	MAPE, RMSE, and Relative Root mean squared error	
[10]	Conducted a study of recently developed forecasting models and predicts the number of confirmed, recovered, and death cases in India caused by COVID-19.	Correlation coefficients and multiple linear regression	R^2 score, RMSE	RMSE: 3085.4305, R^2 Score: 0.9992
[11]	The study area corresponds to the variant identified as Totonaco de la Costa and that identifies the Totonacapan Zone from the political and administrative integration of the municipalities of Mexico	Data Analysis	Graph	
[12]	In this study, data mining models were developed for the prediction of COVID-19 infected patients' recovery using epidemiological dataset of COVID-19 patients of South Korea.	Decision tree, support vector machine, naïve Bayes, logistic regression, random forest, and K-nearest neighbor algorithms	From graphs	
[13]	Making effective predictions for Covid 19 using LSTM-Markov model.	LSTM-Markov model, LSTM	RMSE and R^2 score	Max R^2 Score: 0.88
[14]	LSTM for the prediction of COVID-19 spread in the GCC countries at the Gulf areas: Saudi Arabia, United Arab Emirates, Kuwait, Bahrain, Oman, and Qatar	LSTM	Mean absolute relative error and RMSE	Max RMSE: 1768.35

This paper is organized into the following sections. Section 2 is the literature review, Section 3 provides the dataset details, Section 4 is the proposed method, Section 5 is the evaluation metrics. Section 6 exhibits the results and the comparative analysis. Finally, the conclusion is presented.

II. Literature Review

Table I gives the summary of the previous work on predicting Covid-19 cases. Papers used models such as LSTM (mainly), Logistic regression, decision tree, Support Vector Machine, etc. The evaluation metrics used were RMSE (mainly), R^2 score, Correlation Coefficient, etc. From these results a comparison is done with the model proposed in this paper, given in section VII.

III. Dataset

The dataset used in this work was taken from WHO official website [15]. The tuples of the dataset get updated every day based on the cases for each day. The entire dataset had 200,502 rows and 8 columns. This corresponds to data belonging to 237 countries. From this dataset, the data of the countries under study are then filtered out. This resulted in a dataset having 800 tuples for each country. The dataset of each country had cases from Jan 2020 till April 2022. The initial dataset has the following attributes:

- Date_reported: this attribute represents when the cases are reported
- Country_code: this code is a unique number for each country at the international level
- Country: this is the country name
- WHO_region: these are unique regions recognized by the WHO
- New_cases: represents the new COVID-19 cases reported
- Cumulative_cases: represents the cumulative covid cases from the initial date reported
- New_deaths: represents the new deaths due to Covid-19
- Cumulative_deaths: represents the cumulative deaths from the initial date reported

For further analysis, this study included only one column, i.e., New_cases.

IV. Proposed Method

Figure 1 explains the architecture diagram of our model. First, the dataset pertaining to a country is preprocessed to improve its quality and prepare it for further analysis. Next, the dataset is split into test and train. Out of the total tuples in the dataset, 75% was used for training and 25% was used for testing the model. The train set was used to train the LSTM model and the test set was used to test the LSTM model. The LSTM model was trained using Mean Squared Error (MSE) loss during the epochs of the model. The prediction accuracy of the trained model is measured using correlation coefficient and R^2 score. This process is repeated for all the twelve countries under consideration.

A. Data Preprocessing

The initial step done in this work is data preprocessing. The raw data is preprocessed as follows

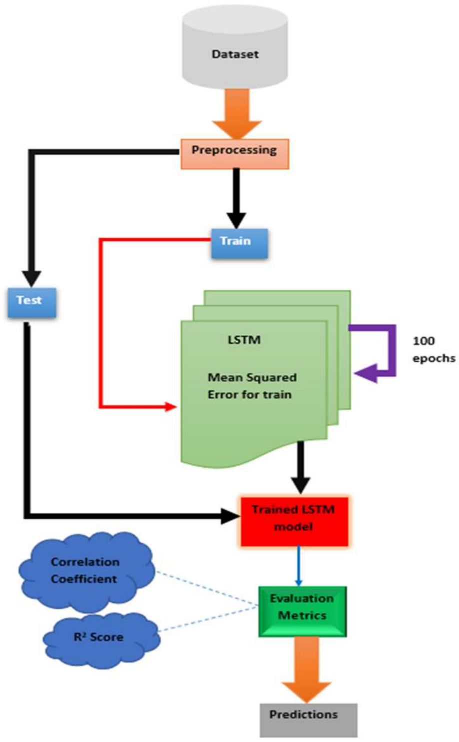

Figure 1: Architecture Diagram

- Remove records that have 0 as a value in the attribute of interest.
- Remove all records that have missing values, i.e., NaN values.
- Perform min-max normalization to the data

B. Prediction Model

This work uses a LSTM model for predicting the Covid-19 cases. LSTM (LSTM) is a type of Recurrent Neural Network (RNNs). It is a time series supervised regression model used in deep learning-based studies. It has feedback connections. It can more precisely forecast the outputs by storing the past in hidden states. It has been used in many applications, such as stock market prediction, weather predictions, to name a few. The LSTM network is composed of cell having three different gates like the following:

A) Input gate
B) Output gate
C) Forgot gate

The cell of LSTM acts as the memory bank of the system. The input gate is where the input is provided to the cell, and it measures the importance of the given input. The output gate is from where the output of the cell is given to the next timestep. The forgot gate not used in this analysis is used to determine if previous information is to be retained or not.

Let, W_i be the weight matrix associated with the input gate i of a cell, b_i be the bias associated with input gate i of a cell, W_o be the weight matrix associated with the output gate o of a cell, b_o be the bias associated with output gate o of a cell, W_c be the weight matrix associated with cell, b_c be the bias at the cell, f_t be the activation at the forgot gate at time t, i_t be the activation at input gate at time t, C_t be the cell state at time t, C_{t-1} be the cell state at time $t-1$, o_t be the activation at the output gate at time t, $\overline{C}_t$ be the candidates for the cell at time t, h_{t-1} be the hidden state of the previous cell or previous timestamp, H_t be the hidden state of the current cell at time t and x_t be the input at time t. The equations at all the gates are given from equations (1) to (6).

$$\mathrm{i_t} = \mathrm{sigmoid}(\mathrm{W_i}[\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_i}) \quad (1)$$

$$o_t = \mathrm{sigmoid}(\mathrm{W_o}[\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_o}) \quad (2)$$

$$H_t = o_t * \tanh(C_t) \quad (3)$$

$$\overline{C_t} = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \overline{C_t} \quad (5)$$

$$Output = Relu(H_t) \quad (6)$$

In this work, we use an LSTM model with two layers. The first LSTM layer has 100 units followed by a dropout of 0.4. The second LSTM layer has 100 units followed by a dropout of 0.2. Finally, a dense layer with RELU activation is used. The rectified linear activation function, or ReLU for short, is a piecewise linear function that, if the input is positive, outputs the input directly; else, it outputs zero. The equation for ReLU function is given in (7).

$$f(x) = \max(0, x) \quad (7)$$

Hence, equation (6) becomes

$$Output = max(0, H_t) \quad (8)$$

$$Output = \begin{cases} 0, & if\ H_t < 0 \\ H_t, & if\ H_t \geq 0 \end{cases} \quad (9)$$

C. Model Parameters

The "learning rate," or the amount of change to the model during each stage of the search process, is known as the "step size," and it is likely the most critical hyperparameter to set for the neural network to obtain optimal performance on your challenge, the value of this hyperparameter has a small positive value, often in the range between 0.00 to 1.00. Reduce LR refers to reducing the learning rate when a metric is no longer improving. When learning becomes stagnant, models often benefit from slowing the learning rate by a factor of 2-10. This callback monitors a quantity and if no improvement is seen for a 'patience' i.e., number of epochs, the learning rate is reduced. This is the learning rate which is used in this research. A patience value of three is used in this research.

Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first order and second-order moments. Adam is a replacement optimization algorithm for stochastic gradient descent for training deep learning models. Adam is the best among the adaptive optimizers in most of the cases.

This study trained the model for 100 epochs and used a batch size of 100. Experiments were conducted by increasing the number of LSTM layers to see if it increases the accuracy. From the results we found out that the model with two layers was sufficient, and the accuracy didn't improve by adding more layers.

V. EVALUATION METRICS

A. Correlation Coefficient

It is an estimator that measures the correlation between two attributes. It is always between -1 and 1, both inclusive. The formula is given in (8).

$$\rho = \frac{\mathrm{Cov(y\_actual, y\_predicted)}}{\sigma_{y\_actual}\, \sigma_{y\_predicted}} \quad (8)$$

where, Cov(y_actual, y_predicted) is the covariance between the actual Covid-19 cases and the predicted Covid-19 cases and the equation is given in (9).

$$\mathrm{Cov}\,(y\_actual, y\_predicted) = \frac{1}{N-1}\sum_{i=1}^{N}(y\_actual_i - \overline{y\_actual})(y\_predicted_i - \overline{y\_predcited}) \quad (9)$$

N is the number of samples for both y_actual and y_predicted, σ_{y_actual} is the standard deviation of y_actual and $\sigma_{y\_predicted}$ is the standard deviation of y_predicted.
If correlation coefficient is near 1, then the prediction made by the model is positively correlated with the actual values. If correlation coefficient is near -1 then the prediction made by the model is negatively correlated with the actual values. If correlation coefficient is near 0 then they are not correlated. In this study it is used as an accuracy metric.

B. R^2 Score

It is another metric used to test the correlation between two attributes or to find the similarity between actual and predicted values. The value ranges from 0-1. However, negative values can occur. In those cases, the model doesn't correlate, or the attributes are not similar. The formula is given in equation (10).

$$R^2 = 1 - \frac{SSR}{TSS} \quad (10)$$

where, SSR is the sum of squares of residuals and TSS is the total sum of squares.

VI. RESULTS

TABLE II. RESULTS FOR LSTM PREDICTION FOR 12 COUNTRIES

Sr No.	Country	Correlation Coefficient		R^2 Score	
		Train	*Test*	*Train*	*Test*
1	USA	0.872283	0.913634	0.759391	0.823396
2	China	0.974068	0.915556	0.943244	0.803036
3	UAE	0.986059	0.983271	0.972035	0.966607
4	India	0.994165	0.994153	0.987467	0.987725
5	Brazil	0.824111	0.803136	0.677774	0.644466
6	France	0.864579	0.806594	0.746944	0.643915
7	Germany	0.931834	0.930201	0.867814	0.860519
8	Spain	0.935426	0.915418	0.874105	0.812801
9	Republic of Korea	0.957758	0.954356	0.917607	0.896829
10	Italy	0.910150	0.933998	0.827837	0.871581
11	Singapore	0.918045	0.940978	0.842595	0.885198
12	Argentina	0.919088	0.983751	0.841942	0.945346

The prediction results for the 12 countries are given in Table 1. From Table 1 the following inferences can be made:

- USA has a train correlation of 0.872283 and a test correlation of 0.913634, while a train R^2 score of 0.759391 and test R^2 score of 0.823396. This signifies that the results of the LSTM predictions are accurate, as the train and test values for both metrics are similar and are highly valued.
- China has a train correlation of 0.974068 and a test correlation of 0.915556, while a train R^2 score of 0.943244 and test R^2 score of 0.803036. The LSTM predictions are accurate, as the train and test values for both metrics are similar and are strongly correlated.
- UAE has a train correlation of 0.986059 and a test correlation of 0.983271, while a train R^2 score of 0.972035 and test R^2 score of 0.966607. This LSTM predictions are highly accurate, as the train and test values for both metrics are highly and strongly correlated with values near to 1.
- India has a train correlation of 0.994165 and a test correlation of 0.994153, while a train R^2 score of 0.987467 and test R^2 score of 0.987725. This LSTM predictions has the highest train and test values for both metrics amongst all countries for this study.
- Brazil has a train correlation of 0.824111 and a test correlation of 0.803136, while a train R^2 score of 0.677774 and test R^2 score of 0.644466. This signifies that the results of the LSTM predictions are loosely correlated, as the train and test values for both metrics are similar and has the lowest amongst all countries in this study.
- France has a train correlation of 0.864579 and a test correlation of 0.806594, while a train R^2 score of 0.746944 and test R^2 score of 0.643915. The LSTM predictions are loosely correlated, as the train and test values are similar and have low values for both metrics.
- Germany has a train correlation of 0.931834 and a test correlation of 0.930201, while a train R^2 score of 0.867814 and test R^2 score of 0.860519. The LSTM predictions are accurate, as the train and test values for both metrics are similar and have high values.
- Spain has a train correlation of 0.935426 and a test correlation of 0.915418, while a train R^2 score of 0.874105 and test R^2 score of 0.812801. The LSTM predictions follow the same trend with both train and test values having similar and peak values for both metrics.
- Republic of Korea has a train correlation of 0.957758 and a test correlation of 0.954356, while a train R^2 score of 0.917607 and test R^2 score of 0.896829. The predictions are highly correlated with both the train and test having peak values.
- Italy has a train correlation of 0.910150 and a test correlation of 0.933998, while a train R^2 score of 0.827837 and test R^2 score of 0.871581. The predictions are accurate, as the train and test values for both metrics are similar and are highly valued.
- Singapore has a train correlation of 0.918045 and a test correlation of 0.940978, while a train R^2 score of 0.842595 and test R^2 score of 0.885198. This signifies that the results of the LSTM predictions are accurate, as the train and test values for both metrics are similar and have strongly correlated values.
- Argentina has a train correlation of 0.919088 and a test correlation of 0.983751, while a train R^2 score of 0.841942 and test R^2 score of 0.945346. The predictions are accurate and are highly valued.

VII. Comparative Analysis

From the literature review, given in Table I, we see that 3 papers are relevant for comparison with the research work carried out in this paper. Table II provides the results obtained in this study.

From Table 1 we see that the maximum correlation coefficient obtained by the model proposed in [1] is 0.76. While the model proposed in this paper has a maximum correlation coefficient of 0.994165, making our model better compared to the model proposed in [1].

From Table 1, the maximum R^2 score obtained by [10] is 0.9992. While this paper's model has a maximum R^2 score of 0.987725 making our model comparatively like the model proposed in [10].

From Table 1, the maximum R^2 score of 0.882 was obtained by the model in [13]. As given in Table 2, this paper's model has a higher R^2 score making the proposed model better than the model in [13].

Conclusion

In this paper, we have used LSTM deep learning algorithm to predict the number of COVID-19 cases in 12 countries using past data. R^2 score and correlation coefficient were used as evaluation metrics. The model went through 100 epochs of training for each country. The parameter used was mean squared error. Different models were made for different countries due to different policies and rules for each of them.

The novelty of this paper is that it is predicting the Covid cases for multiple countries and it is using correlation coefficient as an accuracy metric, rather than a metric of finding the correlation between two variables. The prediction performance of the model was evaluated and found to be appropriate for most of the countries, with little overfitting. Most countries had high R^2 score and correlation coefficient between the train and test datasets. Others had similar R^2 score and correlation coefficient between the train and test. Hence, the predictions are accurate.

Limitation of the study include dataset having missed or extreme values and not having sufficient literature papers having the same metric for the given model. Further, this work can be extended using hybrid models for improving the accuracy of prediction.

References

[1] Yuanyuan Peng, Cuilian Li, et al. "Real-time Prediction of the Daily Incidence of COVID-19 in 215 Countries and Territories Using Machine Learning: Model Development and Validation", 2021, vol 23, iss 6, pp 1-12.

[2] Claudia Barrı́a-Sandoval, Guillermo Ferreira, "Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study", Apr 29th, 2021, pp 1-16.

[3] Hongwei Zhao, Naveed N. Merchant, et al., "COVID-19: Short term prediction model using daily incidence data", PLOS ONE, Apr 14th, 2021, pp 1-14.

[4] Lixiang Li, Zihang Yang, Zhongkai Dang, Cui Meng, Jingze Huang, Haotian Meng, Deyu Wang, Guanhua Chen, Jiaxuan Zhang, Haipeng Peng, Yiming Shao, "Propagation analysis and prediction of the COVID-19", Infectious Disease Modelling, vol. 5, Mar 27th, 2020, pp 282-292.

[5] Munish Kumar, Surbhi Gupta, Krishan Kumar, and Monika Sachdeva. "Spreading of COVID-19 in INDIA, ITALY, JAPAN, SPAIN, UK, US: A Prediction Using ARIMA and LSTM Model", Oct 2020, vol 1, iss 4, pp 1-9, Available [Online]: https://doi.org/10.1145/3411760 .

[6] Rohitash Chandra, Ayush Jain, Divyanshu Singh Chauhan. "Deep learning via LSTM models for COVID-19 infection forecasting in India", Jan 28th, 2022, vol 17, pp 1-28. Available [Online]: https://doi.org/10.1371/journal.pone.0262708 .

[7] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). "Long Short-term Memory", Dec, vol 9, pp 1735-1780, Available [Online]: http://dx.doi.org/10.1162/neco.1997.9.8.1735 .

[8] Annwesha Banerjee Majumder, Somsubhra Gupta, Dharmpal Singh and Sourav Majumder, "An Intelligent System for Prediction of COVID-19 Case using Machine Learning Framework-Logistic Regression", Feb 2021, pages 9, Available [Online]: https://10.1088/1742-6596/1797/1/012011 .

[9] Zuhaira M. Zain, Nazik M. Alturki, "COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach", Journal of Control Science and Engineering, vol. 2021, Article ID 8785636, 23 pages, 2021. https://doi.org/10.1155/2021/8785636 .

[10] Rajani Kumari, Sandeep Kumar, Ramesh Chandra Poonia, Vijander Singh, Linesh Raja. "Analysis and Predictions of Spread, Recovery, and Death Caused by COVID-19 in India", Jun 2021, vol 4 iss 2, pp 65-75, Available [Online]: https://doi.org/10.26599/BDMA.2020.9020013 .

[11] Medel-Ramírez, Carlos and Medel-López, Hilario. "Data mining for the study of the Epidemic (SARS-CoV-2) COVID-19: Algorithm for the identification of patients speaking the native language in the Totonacapan area – Mexico", Aug 2020-Nov 2021, Available [Online]: https://ideas.repec.org/p/pra/mprapa/102039.html .

[12] Sunidhi Shrivastava, Pankaj Gugnani. "Data Mining Prediction Models for Novel Coronavirus (COVID-19) using Machine Learning", Jan 3rd, 2021, Available [Online]: https://ssrn.com/abstract=3759263 .

[13] Ruifang Ma, Xinqi Zheng, Peipei Wang, Haiyan Liu, and Chunxiao Zhang, "The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method", Aug 31st, 2021, Available [Online]: https://doi.org/10.1038/s41598-021-97037-5.

[14] Kareem Kamal A. Ghany, Hossam M. Zawbaa, Heba M. Sabri, "COVID-19 prediction using LSTM algorithm: GCC case study", Mar 31st, 2021, Available [Online]: https://doi.org/10.1016/j.imu.2021.100566

[15] WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/table.