USING FUZZY TIME SERIES FORECASTING AND GAUSSIAN MIXTURE MODEL TO CLASSIFY AND PREDICT NEW CASES OF COVID-19 IN SAUDI ARABIA

by

Safar M. ALGHAMDI[a], Sharaf Obaid ALI[b], Maha A ALDAHLAN[c], Gaafar Mohamed ABDALKRIM[d], and Azhari A. ALHAG[a*]

[a] Department of Mathematics, College of Science, Taif University, Taif, Saudi Arabia
[b] Department of Mathematics and Statistics, College of Sciences, Shaqra University, Kingdom of Saudi Arabia, previously College of Computer Science, Alzaeim Alazhari University, Khartoum, Sudan
[c] Department of Statistics, College of Science, University of Jeddah, Jeddah, Saudi Arabia
[d] College of Science and Humanity, Prince Sattam bin Abdulaziz University, Alsulail, Saudi Arabia

Original scientific paper
https://doi.org/10.2298/TSCI22S1261A

In light of the global events resulting from the spread of the Corona pandemic and viral mutations, there is a need to examine epidemic data in terms of numbers of infected and deaths, different geographical locations, and the dynamics of disease dissemination virus. In the Kingdom of Saudi Arabia (KSA), since the spread of the virus on March 2, 2020, the number of confirmed cases has increased to 599044 cases until January 13, 2022, of which 262 are critical cases, while the number of recovery cases have reached 55035 cases, and deaths are 8901. It is a serious disease, and its spread is difficult to contain. The number of cases has continued to grow rapidly since the first cases appeared. Guess and Buck's model for forecasting time-series data is an important figure that cannot be crossed when predicting fuzzy time-series, although several modifications have been made to the model to improve the accuracy of its results. The Gaussian mixture model and the fuzzy method for modelling new cases in Saudi Arabia were used as machine learning methods to classify and predict new cases of the virus in Saudi Arabia. Foggy time series forecasting. The studied datasets from the World Health Organization from May 15 to August 12, 2020 were used.

Key words: *Gaussian mixture model, fuzzy, machine learning, predict, classify*

Introduction

The Coronavirus has caused global concern since it emerged in China in 2019. The outbreak of this virus has caused an unprecedented global health and humanitarian crisis with widespread social and economic disruptions around the world [1]. As of December 1, 2021, 890266 people died and confirmed cases were 52732967 over the year, which is more than the population of South Korea [2]. According to information issued by the World Bank, global economic losses are estimated at 15 trillion dollars due to the Coronavirus [3]. The economy of China, whose first confirmed case of the Covid-19 virus due on December 1, 2019, in Wuhan, was affected, as the repercussions of this crisis are worse and more influential compared to previous crises, as the growth rate decreased by –3% in 2020, compared to the four previous forecasts for the epidemic where it was Growth in global GDP is expected to rise by 3.3%, and this decline is expected to continue until 2022 when it marked the deepest decline since the Great Depression in nearly a century. The reper-

[*] Corresponding author, e-mail: a.alhag@tu.edu.sa

cussions of the spread of the virus led to a decline in growth rates, high unemployment rates, and the accumulation of Chinese debt. In this paper, we apply the fuzzy time series forecasting and Gaussian mixture model to classify and predict new cases of the virus in Saudi Arabia.

Review of related work

The study [4] sheds light on the coronavirus pandemic, extracted some statistical indicators, and used ANN models and logistic regression models. In [5], the purpose of this study is to investigate the usage of a decision tree approach to predict student academic achievement. Education serves as a platform for society to improve the quality of its population. Improving educational quality necessitates the ability to forecast student success. When developing a decision tree structure, it is used to achieve automatic chi-square interaction detection (CHAID). Financial position, learning motivation, and gender was discovered to influence the student performance. This article [6] used support vector machines to predict the student's future performance in computer science and mathematics based on past performance in computer science, mathematics, and statistics. Students' previous subjects were assessed using modern feature selection techniques to determine any relationship between academic excellence in a particular discipline and prior knowledge of the subject. A classification accuracy of 80% has been achieved using support vector machines, indicating that this method can be designed to create recommendation or orientation systems for students, however, the classification model will still benefit from more teaching examples. The results of this study confirm the potential and benefits of using machine learning techniques to improve teaching and learning in higher education. The Bayesian classification method has also been used by [7]. In the work that was about predicting students' grades based on the previous year's performance. In which the researchers concluded that the study helps the teacher and the student to improve the student's grades. This paper [8] provided an empirical evaluation of several time series models for predicting COVID-19 cases, recovery, and mortality in KSA. Models were trained using self-regressive integrated moving average, exponential smoothing, cube slice, Holt simple exponential smoothing, and Holt winters. Experimental results show that the ARIMA model has a smaller prediction error in the prediction of confirmed cases, which is consistent with the findings in the literature, while the cube chart showed better predictions of recoveries and deaths. As more data becomes available, fluctuations in predictive accuracy measures are observed, possibly due to abrupt changes in data. Also [9], Represents COVID-19 case prediction using the Type 2 fuzzy logic system with Takagi-Sugeno-Kanga fuzzy inference and the neural network learning models. The model parameters were adapted using an inverse gradient ratio approach. It has been established that the proposed model outperforms Type 1 fuzzy logic [10], the system, and the ANN in terms of root mean square error, mean absolute error in percent and mean absolute error. The entire world has been confronted with an unprecedented human challenge as a result of COVID-19, which is caused by a novel coronavirus discovered in 2019, (SARS-CoV-2) [11]. After claiming hundreds of thousands of lives, this virus still has a significant grip on millions of individuals. This virus is highly infectious, with R0 as high as 6.5 worldwide and between 1.5 and 2.6 in India. As a result, the total number of infections and deaths will rise on a daily basis until the curve flattens. Under the current circumstances, it is unavoidable to create a model that can predict future morbidities, recoveries, and deaths [12, 13]. Convolutional Neural Network-based models [14] (one block VGG, two block VGG, three block VGG, four block VGG, LetNet-5, AlexNet, and Resnet-50) were used in this study to detect Coronavirus and SARS MERS infected patients and distinguish them from healthy subjects using lung X-ray scans, which proved to be a difficult task due to overlapping characteristics of different Coronavirus types. Furthermore, in Italy, the LSTM model was employed for time series forecasting of nCOV cases over the next ten days.

When compared to other models, the VGG1 model separates the three classes with an accuracy of about 91%, whilst the technique based on the LSTM forecasts the number of nCOV cases with 99% accuracy in [15]. They developed several models based on ARIMA and FUZZY time series methodology to predict infection, mortality, and recovery from COVID-19 in India and Maharashtra, which is the most affected state in India, explicitly tracking statistics COVID-19 before Lockdown 3.0 (May 17, 2020). They showed that both models assume an exponential increase in the number of COVID-19 cases soon, the COVID-19 dataset was predicted for the next seven days, and they acknowledge that the predicted values are in good agreement with the true ones. the implications of all six COVID-19 scenarios for Maharashtra and India. They concluded that the forecasts for the ARIMA and FUZZY time series models would be useful for policymakers in health systems so that the system and medical staff can prepare to deal with a pandemic [16].

Gaussian mixture model

A clustering approach known as the Gaussian mixture model (GMM) is employed to identify the underlying groups of data [17]. It can be viewed as a probabilistic model where each group's means and covariances serve as the parameters and Gaussian distributions are assumed for each group. The total number of weighted Gaussians makes up a GMM. A new symbol π is introduced to represent the weights. For example, πk means the probability that x belongs to the k^{th} Gaussian:

$$p(x)=\pi_1 N\left(\mu_1 \sum{}_1\right)+\pi_2 N\left(\mu_2 \sum{}_2\right)+\pi_3 N\left(\mu_3 \sum{}_3\right) \tag{1}$$

where $\sum_i$, i = 1, 2, 3 is the covariance matrix. Note that there is a limitation – the sum of all π must be equal to 1. To prove this, we use the fact that all probability distributions must be integrated to 1. The $P(x)$ is integrated to 1, and each Gaussian (since we know that it is real distributions) can also be integrated to 1. Thus, the sum of all π is equal to 1:

$$\begin{gathered}1=\int p(x)\mathrm{d}x=\int \pi_1 N\left(\left(\frac{x}{\mu_1}\sum{}_1\right)\mathrm{d}x+\pi_2 N\left(\left(\frac{x}{\mu_2},\pi_1 N\left(\frac{x}{\mu_1}\right)\mathrm{d}x\right.\right.\right.\\ 1=\pi_1\times 1+\pi_2\times 1\end{gathered} \tag{2}$$

Another way to represent it is to introduce a new hidden variable Z. The Z shows what data a given Gaussian represents. You can write:

$$\pi_k=P(Z=k) \tag{3}$$

It turns out that there is some hidden variable Z, which we do not know and which we cannot measure, but each of these Z causes the appearance of a Gaussian, and from our data we can see the total effect of all individual Z. This is important because it puts Gaussian mixtures distributions in the framework of maximizing expectations.

Fuzzy

Song and Chissom [18] were among the first scientists, whose study such problems and they had propose the time series fuzzy models in 1993. They proposed some concepts of fuzzy time series and its models. A big snag of the FTS models which developed by Chissom and Song [18] were that they are associated with unnecessary high computational overheads due to complex matrix operations, in order to reduce the computational overhead of the time-variant and time-invariant models where presented a simple study in which he tried to avoid the computational complications in the model, Chen's model have the following steps:

– The first step is to partition the universe of the discourse into equal lengthy intervals.

– The second step is to define the fuzzy sets on the universe of the discourse.
– The third Step is to obfuscate the historical data.
– The fourth Step is to identify the fuzzy relationships (FLR).
– The fifth step is to establish the fuzzy relationship groups (FLRG).
– The last step is to defuzzify the forecasted output.

Define the universe of discourse and partition it into equally lengthy intervals

Definition: Let $Y(t)\varepsilon R$, $(t = 0, 1, 2,...)$ be the time series, if $f_i(t)$ is the fuzzy set in $Y(t)$ and $F(t) = \{f_1(t), f_2(t),...\}$ then $F(t)$ is called the fuzzy time series in $Y(t)$. It is noted that $F(t)$ can be regarded as the linguistic variable and $f_i(t)(i = 0,1,2,..., m)$ can be viewed as possible linguistic [18].

Define fuzzy sets on the universe of discourse

Let $A_1, A_2,..., A_k$ to be fuzzy sets which are linguistic values of the linguistic variable new cases of COVID-19 in KSA:

$$A = f_A : U \to \frac{[0,1](u_i)}{u_1} + \frac{f_A(u_2)}{u_2} + \ldots + \frac{f_A(u_K)}{u_K} \tag{4}$$

where the universe of discourse U: $U = \{u_1, u_2,..., u_k\}$ is defined as $[D_{\min} - D_1, D_{\max} - D_2]$. Such that $D_1 = 57$ and $D_2 = 481$. The $D_{\min}$ and $D_{\max}$, we partition the universe of discourse U into seven evenly lengthy $U = [1200, 5400]$, A_1: [1200-1800], A_2: [1200-2400], A_3: [2400-3000], A_4: [3000-3600], A_5: [3600-4200], A_6: [4200-4800], A_7: [4800-5400].

The linguistic values of the linguistic variable new cases of Covid-19 in KSA, are defined on the universe of discourse:

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}, \quad A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}$$

$$A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5} + \frac{0}{u_6} + \frac{0}{u_7}, \quad A_5 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6} + \frac{0}{u_7}$$

$$A_6 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7}, \quad A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7}$$

$$U = [1200\text{-}5400] \text{ such that, } A_1 : [1200\text{-}1800],\ A_2 : [1800\text{-}2400],\ A_3 : [2400\text{-}3000]$$

$$A_4 : [3000\text{-}3600]\, A_5 : [3600\text{-}4200],\ A_6 : [4200\text{-}4800],\ A_7 : [4800\text{-}5400]$$

Identify fuzzy relationships

According to Song and Chen's work [19], they assumed that if there exists the fuzzy relationship $R(t-1 \to t)$ such that $F(t-1) \to F(t) \times R(t-1 \to t)$ where $\times$ represent an operator, then $F(t)$ is said to be caused by$F(t-1)$, this relation called first order time series and denoted by the expression $F(t-1) \to F(t)$.

Fuzzify historical data

In this context, fuzzification is the process of identifying associations between the historical values in the dataset and the fuzzy sets defined in the previous step. Each historical value is fuzzified according to its highest degree of membership. If the maximum member ship of one day's the number of new cases of the disease, say $F(t-1)$ occurs at fuzzy set A_j then the fuzzified number of new cases at of the disease $F(t-1)$ treated as A_j, in this paper the following linguistic variable were used, A_1 represent very low level of new cases recorded, A_2 low level, A_3 changeless level, A_4 moderate, A_5 normal level, A_6 high level, A_7 very high of new cases recorded.

Table 1. Fuzzification of historical data for actual of new cases (NC) and Fuzziefied new cases (FNC), each historical value is fuzzified according to its highest degree of membership

Date	NC	Interval	FNC	date	NC	interval	FNC
2020-05-15	2039	[1800-2400]	A_2	2020-06-29	3989	[3600-4200]	A_5
2020-05-16	2307	[2400-3000]	A_3	2020-06-30	3943	[3600-4200]	A_5
2020-05-17	2840	[2400-3000]	A_3	2020-07-01	4387	[4200-4800]	A_6
2020-05-18	2736	[2400-3000]	A_3	2020-07-02	3402	[3000-3600]	A_4
2020-05-19	2593	[2400-3000]	A_3	2020-07-03	3383	[3000-3600]	A_4
2020-05-20	2509	[2400-3000]	A_3	2020-07-04	4193	[3600-4200]	A_5
2020-05-21	2691	[2400-3000]	A_3	2020-07-05	4128	[3600-4200]	A_5
2020-05-22	2532	[2400-3000]	A_3	2020-07-06	3580	[3000-3600]	A_4
2020-05-23	2642	[2400-3000]	A_3	2020-07-07	4207	[4200-4800]	A_6
2020-05-24	2442	[2400-3000]	A_3	2020-07-08	3392	[3000-3600]	A_4
2020-05-25	2399	[1800-2400]	A_2	2020-07-09	3036	[3000-3600]	A_4
2020-05-26	2235	[1800-2400]	A_2	2020-07-10	3183	[3000-3600]	A_4
2020-05-27	1931	[1800-2400]	A_2	2020-07-11	3159	[3000-3600]	A_4
2020-05-28	1815	[1800-2400]	A_2	2020-07-12	2994	[2400-3000]	A_3
2020-05-29	1644	[1200-1800]	A_1	2020-07-13	2779	[2400-3000]	A_3
2020-05-30	1581	[1200-1800]	A_1	2020-07-14	2852	[2400-3000]	A_3
2020-05-31	1618	[1200-1800]	A_1	2020-07-15	2692	[2400-3000]	A_3
2020-06-01	1877	[1800-2400]	A_2	2020-07-16	2671	[2400-3000]	A_3
2020-06-02	1881	[1800-2400]	A_2	2020-07-17	2764	[2400-3000]	A_3
2020-06-03	1869	[1800-2400]	A_2	2020-07-18	2613	[2400-3000]	A_3
2020-06-04	2171	[1800-2400]	A_2	2020-07-19	2565	[2400-3000]	A_3
2020-06-05	1975	[1800-2400]	A_2	2020-07-20	2504	[2400-3000]	A_3
2020-06-06	2591	[2400-3000]	A_3	2020-07-21	2429	[2400-3000]	A_3
2020-06-07	3121	[3000-3600]	A_4	2020-07-22	2476	[2400-3000]	A_3
2020-06-08	3045	[3000-3600]	A_4	2020-07-23	2331	[1200-2400]	A_2
2020-06-09	3369	[3000-3600]	A_4	2020-07-24	2238	[1200-2400]	A_2
2020-06-10	3288	[3000-3600]	A_4	2020-07-25	2378	[1200-2400]	A_2
2020-06-11	3717	[3600-4200]	A_5	2020-07-26	2201	[1200-2400]	A_2
2020-06-12	3733	[3600-4200]	A_5	2020-07-27	1968	[1200-2400]	A_2
2020-06-13	3921	[3600-4200]	A_5	2020-07-28	1993	[1200-2400]	A_2
2020-06-14	3366	[3600-4200]	A_4	2020-07-29	1897	[1200-2400]	A_2
2020-06-15	4233	[3600-4200]	A_6	2020-07-30	1759	[1200-1800]	A_1
2020-06-16	4507	[3600-4800]	A_6	2020-07-31	1629	[1200-1800]	A_1
2020-06-17	4267	[3600-4800]	A_6	2020-08-01	1686	[1200-1800]	A_1
2020-06-18	4919	[3600-4800]	A_6	2020-08-02	1573	[1200-1800]	A_1
2020-06-19	4757	[3600-4800]	A_6	2020-08-03	1357	[1200-1800]	A_1
2020-06-20	4301	[3600-4800]	A_6	2020-08-04	1258	[1200-1800]	A_1
2020-06-21	3941	[3600-4200]	A_5	2020-08-05	1342	[1200-1800]	A_1
2020-06-22	3379	[3000-3600]	A_4	2020-08-06	1389	[1200-1800]	A_1
2020-06-23	3393	[3000-3600]	A_4	2020-08-07	1402	[1200-1800]	A_1
2020-06-24	3139	[3000-3600]	A_4	2020-08-08	1567	[1200-1800]	A_1
2020-06-25	3123	[3000-3600]	A_4	2020-08-09	1469	[1200-1800]	A_1
2020-06-26	3372	[3000-3600]	A_4	2020-08-10	1428	[1200-1800]	A_1
2020-06-27	3938	[3600-4200]	A_5	2020-08-11	1257	[1200-1800]	A_1
2020-06-28	3927	[3600-4200]	A_5	2020-08-12	1521	[1200-1800]	A_1

Identify fuzzy relationships

If the time series variable $F(t)$ is fuzzified as $A_j \rightarrow A_k u_1, u_2, \ldots u_n$ and $F(t)$ as A_k, A_j then is related to A_k, $A_j \rightarrow A_k$. If there exists a one-to-many relationship in the relationship group of A_j, $A_2 \rightarrow A_2, A_1, \ldots A_n$ and the maximum degrees of belongingness occurs at set $u_1, u_2, \ldots u_n$ then the forecasted output is computed as the average of the midpoints.

Table 2. Shows the fuzzy logical relationship for Chen's first-order model for the new cases, actual cases (AC), the forecasted output is computed as the average of the midpoints

Date	AC	Ff	FE	FLRG	Interval mid point
2020-05-15	2039	A_2		$A_2 \rightarrow A_2, A_1, \ldots A_n$	1500, 2100, 2700
2020-05-16	2307	A_3	2100	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-17	2840	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-18	2736	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-19	2593	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-20	2509	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700; 3300
2020-05-21	2691	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100; 2700, 3300
2020-05-22	2532	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-23	2642	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-24	2442	A_3	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-25	2399	A_2	2700	$A_3 \rightarrow A_3, A_2, \ldots A_4$	2100, 2700, 3300
2020-05-26	2235	A_2	2700	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-05-27	1931	A_2	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-05-28	1815	A_2	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-05-29	1644	A_1	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-05-30	1581	A_1	1800	$F(t) = F(t-1) \rightarrow F(t) \times R(t-1 \rightarrow t)$	1500, 2100
2020-05-31	1618	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-06-01	1877	A_2	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-06-02	1881	A_2	1800	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500; 2100; 2700
2020-06-03	1869	A_2	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500; 2100; 2700
2020-06-04	2171	A_2	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-06-05	1975	A_2	2100	$A_2 \rightarrow A_2, A_1, \ldots A_3$	1500, 2100, 2700
2020-06-06	2591	A_3	2100	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-06-07	3121	A_4	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-06-08	3045	A_4	2700	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-09	3369	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-10	3288	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-11	3717	A_5	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-12	3733	A_5	3600	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-13	3921	A_5	3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-14	3366	A_4	3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-15	4233	A_6	3900	$A_6 \rightarrow A_4, A_6, A_6 A_7$	3300, 4500, 5100
2020-06-16	4507	A_6	4300	$A_6 \rightarrow A_4, A_6, A_6 A_7$	3300, 4500, 5100
2020-06-17	4267	A_6	4300	$A_6 \rightarrow A_4, A_6, A_6 A_7$	3300, 4500, 5100
2020-06-18	4919	A_7	4300	$A_7 \rightarrow A_6$	4500, 5100
2020-06-19	4757	A_6	4800	$A_6 \rightarrow A_4, A_6, A_6 A_7$	3300, 4500, 5100
2020-06-20	4301	A_6	4300	$A_6 \rightarrow A_4, A_6, A_6 A_7$	3300, 4500 5100
2020-06-21	3941	A_5	4300	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-22	3379	A_4	3900	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-23	3393	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-24	3139	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-25	3123	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-26	3372	A_5	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-06-27	3938	A_5	3600	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-28	3927	A_5	3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500

Table 3. Fuzzy logical relationships for Chen's first-order model forecasted the number of confirmed cases in Saudia Arabia

Date	AC	AC	FE	FLRG	Interval mid-point
2020-06-29	3989		3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-06-30	3943	A_5	3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-07-01	4387	A_6	3900	$A_6 \rightarrow A_4, A_6, A_7$	3300, 4500, 5100
2020-07-02	3402	A_4	4300	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-03	3383	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-04	4193	A_5	3600	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-07-05	4128	A_5	3900	$A_5 \rightarrow A_4, A_5, A_6$	3300, 3900, 4500
2020-07-06	3580	A_4	3900	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-07	4207	A_6	3600	$A_6 \rightarrow A_4, A_6, A_7$	3300, 4500, 5100
2020-07-08	3392	A_4	4300	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-09	3036	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-10	3183	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-11	3159	A_4	3600	$A_4 \rightarrow A_3, A_4, A_5, A_6$	2700, 3300, 3900, 4500
2020-07-12	2994	A_3	3600	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-13	2779	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-14	2852	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-15	2692	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-16	2671	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-17	2764	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-18	2613	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-19	2565	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-20	2504	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-21	2429	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	22700, 3300
2020-07-22	2476	A_3	2700	$A_3 \rightarrow A_3, A_2, A_4$	2100, 2700, 3300
2020-07-23	2331	A_2	2700	$A_3 \rightarrow A_3, A_2, A_4$	1500, 2100, 2700
2020-07-24	2238	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-25	2378	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-26	2201	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-27	1968	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-28	1993	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-29	1897	A_2	2100	$A_2 \rightarrow A_2, A_1, A_3$	1500, 2100, 2700
2020-07-30	1759	A_1	2100	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-07-31	1629	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-01	1686	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-02	1573	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-03	1357	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-04	1258	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-05	1342	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-06	1389	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-07	1402	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-08	1567	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-09	1469	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-10	1428	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500; 2100
2020-08-11	1257	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100
2020-08-12	1521	A_1	1800	$A_1 \rightarrow A_1, A_2$	1500, 2100

The Gaussian mixture model

Descriptive statistics

Table 4. Descriptive statistics

Std. deviation	Mean	Maximum	Minimum	Observations without missing data	Observations with missing data	Observations	Variable
944.218	2740.422	4919.000	1257.000	90	0	90	New cases

Table 5. Evolution of the BIC for each model

5	4	3	2	Model/number of classes
–1519.802	–1510.803	–1506.237	–1497.2378	Equal variance

Table 6. Proportions, the mean and the variance by class

2	1	Class
0.369	0.631	Proportions
3706.207	2176.856	Mean
337356.763	337356.763	Variance

Table 7. Selection criterion for the selected model

DF	Entropy	NEC	Log-likelihood	ICL	AIC	BIC
4.000	20.119	4.890	–739.619	–1537.476	–1487.239	–1497.238

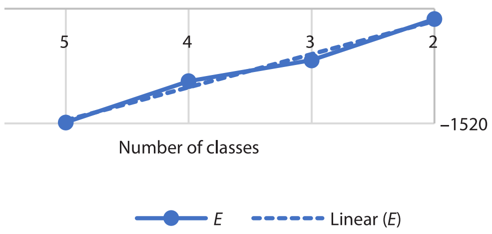

Figure 1. Evolution of the BIC for each model

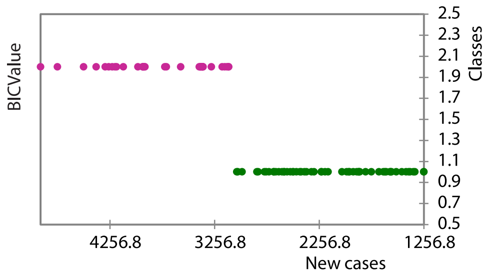

Figure 2. The MAP classification

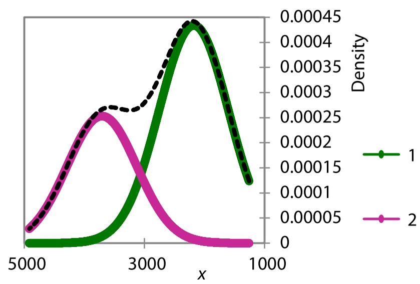

Figure 3. Fitted model

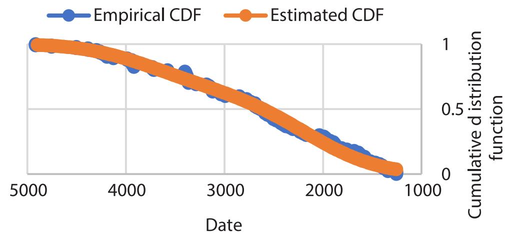

Figure 4. The empirical and estimated (cumulative distribution function)

In this part, we present some of the basic parameters of the data that we used, including the mean and standard deviation. As we see in tab. 4, the mean is 2740.422 and the standard deviation equals 944.218. The evolutions of the BIC for each model are presented in tab. 5. The mean for class one, as shown in tab. 6, is 217.856, for class two, it is 3706.207. For determining the number of clusters produced by a mixture model, NEC performed well, see tab. 7.

As illustrated in fig. 1 each model's BIC has evolved, the lower BIC value indicating lesser penalty terms and, therefore, a better model. Figure 2 shows the results of the covid-19 data's Gaussian mixture modeling. There are exactly two clusters in this segment. The Gaussian mixture model recognized by two clusters in our sample is seen in fig. 3. Each Gaussian in the mixture is made up of its mean, which serves as its center, and its covariance, which serves as its width. As demonstrated in fig. 4, there are no deviations between the estimated CDF and the empirical CDF.

Conclusion

From the observations made regarding the data sets considered during the analysis, it can be concluded that the GMM approach can accurately classify and predict algorithm for a large data set, tab. 6. It should also be noted that based on the result, it was concluded that the two models were accurate and valid. It is also noted that this epidemic is characterized by rapid spread, and the pattern of changes in infection numbers is usually random and takes the nature of random processes, where the focus of infection can be detected, which leads due to the inability to accurately predict the number of infected people. Therefore, the ambiguity of the results in this case is important to give decision makers some room to expect, even in the event of significant changes. In this study, the Chin model was used by replacing the numbers of students enrolled with the numbers of newly registered cases of COVID-19, tabs. 1 and 2. The results were puzzled for the period from 5/15/10/8 and then predicted in this study that the trend pattern The change in the number of new cases for the period from 11/8 to 15/8 was predicted and compared with the actual number of new cases recorded in that period. The results indicated that the number of new cases continued to decline, and that the results of predicting the numbers of people infected with COVID-19 disease occurred in a large proportion with the real numbers for the same period. We note that the actual numbers fall within the expected foggy period, tab. 2, which is – a very low level of new cases recorded – which means a sharp decrease in the number of infected people.

References

[1] Mohamad, A., *et al*., Prediction of Fuzzy Sunspot Time Series by Using RBFANN, *Tikrit Journal of Pure Science, 22* (2018), 11, pp. 106-112
[2] ***, Coronavirus, W. H. O. (2021), *Dashboard*, Mar., 15
[3] ***, International Debt Statistics, https://databank.worldbank.org/source/international-debt-statistics
[4] Elhag, A. A., *et al*., Artificial Neural Networks and Statistical Models for Optimization Studying COVID-19, Results in Physics, *25* (2021), 104274
[5] Kolo, D. K., Solomon, A. A., A Decision Tree Approach for Predicting Students Academic Performance, *Int. J. Education Manag. Eng*., *5* (2015), 5, pp. 12-19
[6] Kenekayoro, P., An Exploratory Study on the Use of Machine Learning to Predict Student Academic Performance, *International Journal of Knowledge-Based Organizations, 8* (2018), 4, pp. 67-79
[7] Bekele, R., Menzel, W., A Bayesian Approach to Predict Performance of a Student (Bapps): A Case with Ethiopian Students, *Algorithms*, *22* (2005), 24
[8] Al-Turaiki, I., *et al*., Empirical Evaluation of Alternative Time-Series Models for Covid-19 Forecasting in Saudi Arabia, *International Journal of Environmental Research and Public Health*, *18* (2021), 8660
[9] Alballa, N., Al-Turaiki, I., Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review, *Informatics in Medicine Unlocked*, *24* (2021), 100564

[10] Zrieq, R., *et al*., Time-Series Analysis and Healthcare Implications of COVID-19 Pandemic in Saudi Arabia, *Healthcare*, *10* (2022), 1874, pp. 1-27
[11] Karnik, N., *et al*., Applications of Type-2 Fuzzy Logic Systems to Forecasting of Time-Series, *Information sciences, 120* (1999) 1, 4, pp. 89-111
[12] Kannan, K., *et al*., A Comparison of Fuzzy Time Series and ARIMA Model, *International Journal of Scientific & Technology Research, 8* (2019), 8, pp. 1872-1876
[13] Chen, S.-M., Forecasting Enrollments Based on Fuzzy Time Series, *Fuzzy Sets and Systems*, *81* (1996), 3, pp. 311-319
[14] Younis, M. C., Evaluation of Deep Learning Approaches for Identification of Different Corona-Virus Species and Time Series Prediction, *Computerized Medical Imaging and Graphics*, *90* (2021), 101921
[15] Huarng, K., Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series, *Fuzzy Sets and Systems., 123* (2001), 3, pp. 87-394
[16] Zhao, H., *et al*., Adaptive Neuro-Fuzzy Inference System for Generation of Diffuser dot Patterns in Light Guides, *Applied Optics*, *49* (2010), 14, pp. 2694-2702
[17] Fraley, C., Raftery, A. E., How Many Clusters, Which Clustering Method, Answers Via Model-Based Cluster Analysis, *The Computer Journal*, *41* (1998), 8, pp. 578-588
[18] Song, Q., Chissom, B. S., Fuzzy Time Series and its Models, *Fuzzy Sets and Systems*, *54* (1993), 1, pp. 1-9
[19] Sah, M., Degtiarev, K. Y., Forecasting Enrollment Model Based on First-Order Fuzzy Time Series, *In World Academy of Science, Engineering and Technology*, *1* (2005), pp. 375-378

Paper submitted: August 20, 2022
Paper revised: October 19, 2022
Paper accepted: November 5, 2022