

## Machine Learning and Classical Forecasting Methods Based Decision Support Systems for COVID-19

Ramazan Ünlü<sup>1</sup> and Ersin Namlı<sup>2,\*</sup>

**Abstract:** From late 2019 to the present day, the coronavirus outbreak tragically affected the whole world and killed tens of thousands of people. Many countries have taken very stringent measures to alleviate the effects of the coronavirus disease 2019 (COVID-19) and are still being implemented. In this study, various machine learning techniques are implemented to predict possible confirmed cases and mortality numbers for the future. According to these models, we have tried to shed light on the future in terms of possible measures to be taken or updating the current measures. Support Vector Machines (SVM), Holt-Winters, Prophet, and Long-Short Term Memory (LSTM) forecasting models are applied to the novel COVID-19 dataset. According to the results, the Prophet model gives the lowest Root Mean Squared Error (RMSE) score compared to the other three models. Besides, according to this model, a projection for the future COVID-19 predictions of Turkey has been drawn and aimed to shape the current measures against the coronavirus.

**Keywords:** Covid-19, machine learning, time series forecasting.

### 1 Introduction

After the appearance of the COVID-19 in December 2019 in Wuhan, China, it quickly spread to almost all countries and ceased to be a problem of China alone. While the world is trying to recognize the COVID-19 virus, an unknown enemy, has to fight it, and it is understood that managing a process in such obscure is as much as fighting the virus is difficult activity.

An all-out struggle with the virus whose epidemiological characteristics are not yet fully known has led to the emergence of economic problems as well as health problems. The rapid and easy transmission of the virus, the lack of a proven treatment process, has brought the healthcare systems of countries to a standstill. Many countries try to maintain economic activities in addition to intensive care capacity, equipment and staff shortages, trying to keep the virus infection under control with restrictions and prohibitions. The timing and degrees of restrictions imposed by states to societies can be cited as a vital factor in controlling the transmission of the virus.

---

<sup>1</sup> Gumushane University, Bağlarbaşı Mah, Gümüşhane, 29000, Turkey.

<sup>2</sup> Istanbul University-Cerrahpasa, İstanbul Üniversitesi Avcılar Kampüsü, İstanbul, 34320, Turkey.

\* Corresponding Author: Ersin Namlı. Email: enamlı@istanbul.edu.tr.

Received: 01 May 2020; Accepted: 24 May 2020.

To be able to provide quality service to patients while optimizing time and limited resources such as medical personnel, medical supplies and protective equipment. is one of the key points of the struggle. Therefore, it is vital to make successful predictions within the scope of the fight against the virus and disease. In this study, the efficient use of limited resources by estimating the number of future infected patients and planning resources for the future in accordance with the number of patients is intended. As a result of estimates, how many patients will have on which date and how much resources will be needed can be calculated. By achieving, resources can be reorganized in a country and can be transferred to another country in which there is a lack of resources. As a result, our work can be described as a fundamental study that forms the infrastructure of a global resource allocation activity.

The number of studies in the literature, from the spread of the COVID-19 virus to the present day, is limited. In particular, the studies in which techniques such as machine learning and deep learning have been applied have remained in the background a little more. Here, we have briefly explained studies related to COVID-19 emphasized how this study can contribute to the literature. Jenny et al. [Jenny, Jenny, Gorji et al. (2020)] studied on six different scenarios. Increasing test numbers and maintaining social distance will decrease the number of infected people and deaths compared to the scenario where there is no mitigation activity is envisaged. The developed model suggests that test strategies have an equal effect with the concept of social distance, but economic costs will be less. Liu et al. [Liu, Magal, Seydi et al. (2020)] showed the effects of implementing major government public policy measures in the model they developed using constant propagation rate in early exponential growth of the COVID-19 epidemic. Rossa et al. [Rossa, Lee, Luo et al. (2020)] used the generalized logistic growth model (GLM), the Richards model and sub-epidemic wave methods that are used for short-term estimation of infectious diseases such as SARS, ebola, pandemic influenza, and dengue in order to estimate the near-future values of COVID-19 case numbers in different provinces in Hubei and China as of February 9, 2020, including 5 days, 10 days and 15 days later.

Funk et al. [Funk, Camacho, Kucharski et al. (2018)] used 2013-2016 West African Ebola epidemic data in their study and combined flexibility with mechanistic models to incorporate uncertainty about epidemic dynamics into the model and presented a model for combating future epidemics. Pirouz et al. [Pirouz, Haghshenas and Piro (2020)] used Group Method of Data Handling (GMDH) algorithm and regression analysis methods to predict approved cases and achieved successful results in their case study with data from Hubei, China. Li et al. [Li, Qin, Xu et al. (2020)] suggested a deep learning model for detection of COVID-19, they have used 4356 volumetric chest CT exams as dataset. Wu et al. [Wu, Leung and Leung (2020)] used data from Dec 31, 2019 to Jan 28, 2020, on the number of cases exported from Wuhan internationally and suggested a Markov Chain Monte Carlo based forecasting model for the potential domestic and international spread of the COVID-19. Allam et al. [Allam and Jones (2020)] pointed universal data sharing in scope of smart city network and benefits of artificial intelligence for pandemic disasters. Dep et al. [Dep and Majumdar (2020)] used auto-regressive integrated moving average (ARIMA) method with time-dependent parameters in order to estimate reproduction number of COVID-19. Randhawa et al. [Randhawa, Soltysiak, El Roz et al.

(2020)] have combined supervised machine learning and digital signal processing for genome analysis, augmented by a decision tree approach to the machine learning component, and a Spearman's rank correlation coefficient analysis to validate the proposed methodology. They have achieved 100% accurate classification of the COVID-19 virus sequences. Barstugan et al. [Barstugan, Ozkaya and Ozturk (2020)] have used 150 CT images to correctly identify the COVID-19 patients. They have applied different feature selection methods and achieve 99.68% accuracy with the 10-fold cross validation techniques and Grey-Level Size Zone Matrix. Jiang et al. [Jiang, Coffee, Bari et al. (2020)] have proposed an artificial intelligence (AI) framework to provide rapid clinical decision-making support. They have used AI with the predictive analytics to estimate severe cases. They have achieved to estimate severe patients with the accuracy rate from 70% to 80%.

In this study, we have aimed to create a prediction model to correctly estimates future of the COVID-19 in top-seven countries in terms of confirmed and death cases. By doing this, the countries can take extra measures against the virus or they can make a better planning for when the measures taken can be loosened.

The remainder of this study is structured as follows. In Section 2, we have briefly explained used forecasting models. In Section 3, we have given the results of each method for the chosen countries. Finally, in Section 4, we have conclude the study with the conclusion and discussion part.

## 2 Methodology

In this study, we have used for different regression model called Support Vector Machines (SVM), Holt-Winters' forecasting, Prophet forecasting, and LSTM. In this section, the methodological foundation of each method is briefly explained.

### 2.1 Support vector machines (SVM)

Support Vector Machines are commonly used for the linearly and non-linearly separable classification problems [Chauhan, Dahiya and Sharma (2019)]. Also, it can be extended for the regression problem and named as Support Vector Regression. In order to understand the methodology behind it, assume we have given a training dataset  $\{(x_1, y_2), \dots, (x_l, y_l)\}$ , where each  $x_i \in R$ , the decision function is given by Eq. (1).

$$f(x) = w\phi(x) + b \quad (1)$$

with respect to  $w_i \in R$  and  $b \in R$ , where  $\phi$  denotes a non-linear mapping from  $R^n$  to higher dimensional space. To ensure  $f(x)$  is as flat as possible, it is needed to find it with the minimum magnitude of weights as shown in Eq. (2).

$$J(w) = \frac{1}{2} \|w\| \quad (2)$$

Subject to all residuals having a value less than  $\varepsilon$ ; or, in equation form (see Eq. (3)):

$$w\phi(x_i) + b - y_i \leq \varepsilon \quad (3)$$

It can be seen that it is not possible to meet this condition for all points. Thus, we can add slack variables  $\xi^+$  and  $\xi^-$  to provide some flexibility and reformulate it as shown below in Eq. (4):

$$J(w) = \frac{1}{2} \|w\|^2 + C \sum_i^n \xi^+ + \xi^- \quad (4a)$$

Subject to: (4b)

$$y_i - (w\phi(x_i) + b) \leq \varepsilon + \xi^+ \quad (4c)$$

$$(w\phi(x_i) + b) - y_i \leq \varepsilon + \xi^- \quad (4d)$$

$$\xi^+ \geq 0 \quad (4e)$$

$$\xi^- \geq 0 \quad (4f)$$

where  $C$  is a constant fixed value that controls the penalty value imposed on the variable which lies outside the  $\varepsilon$  margin and helps to avoid being overfitting.

Ultimately, one can calculate the loss function that ignores the error if the predicted value is less than or equal to  $\varepsilon$ . Thus, it can be formulated as shown in Eq. (5).

$$f(x) = \begin{cases} 0, & \text{if } w\phi(x_i) + b - y_i \leq \varepsilon \\ |w\phi(x_i) + b - y_i| - \varepsilon, & \text{otherwise} \end{cases} \quad (5)$$

For mathematical convenience, the optimization problem described above can be solved in dual form.

## 2.2 Holt-Winters' forecasting

Holt-Winters' forecasting model is developed to capture seasonality effect on time series [Hansun, Charles and Indrati (2019)]. Holt-Winters' forecasting model contains prediction equation and three smoothing functions one for the level  $l_t$ , one for the trend  $b_t$ , and one for the seasonal component  $s_t$ , with the corresponding smoothing parameters  $\alpha$ ,  $\beta^*$ , and  $\gamma$ . And,  $m$  denotes the number of seasons. For example, for the weekly data,  $m$  is equal to 52. There exist two different variations of this method. They are additive and multiplicative. The additive model is used when the seasonal variations are almost constant throughout the series. And, the multiplicative method is used when seasonal variations are changing proportionally to the level of series. In our study, we have used the additive model, which is formulated as in Eqs. (6)-(9).

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (6)$$

$$l_t = \alpha(y_t - s_{(t-m)}) + (1 - \alpha)(l_{(t-1)} + b_{(t-1)}) \quad (7)$$

$$b_t = \beta^*(l_t - l_{(t-1)}) + (1 - \beta^*)b_{(t-1)} \quad (8)$$

$$s_t = \gamma(y_t - l_{(t-1)} - b_{(t-1)}) + (1 - \gamma)s_{(t-m)} \quad (9)$$

where  $k$  is the integer part of  $(h - 1)/m$ , which ensures that estimated of the seasonal indices are coming from the last season of the data sample. The level equation  $l_t$  shows a weighted average between the seasonally adjusted observation and the non-seasonal forecast, which are  $y_t - s_{(t-m)}$  and  $l_{(t-1)} + b_{(t-1)}$ , respectively, for the time  $t$ . The

seasonal formulation shows a weighted average between the current season index,  $y_t - l_{(t-1)} - b_{(t-1)}$ , and the previous season index. The formulation for the seasonal component can be written as in Eq. (10).

$$s_t = \gamma^*(y_t - l_t) + (1 - \gamma^*)s_{(t-m)} \quad (10)$$

If  $l_t$  is substituted from the smoothing equation for the level of the component form, we can get seasonal formulation as shown in Eq. (11).

$$s_t = \gamma^*(1 - \alpha)(y_t - l_{(t-1)} - b_{(t-1)}) + [1 - \gamma^*(1 - \alpha)]s_{(t-m)} \quad (11)$$

Eq. (11) is identical to the seasonal component equation as shown in Eq. (9) with  $\gamma = \gamma^*(1 - \alpha)$ . From this,  $\gamma$  will be greater than or equal to 0 and less than or equal to  $1 - \alpha$ .

### 2.3 Prophet forecasting

The prophet forecasting model is based on the idea of fitting the Generalized Additive model. The prophet is published by Facebook's Core Data Science team and the main study can be found in Taylor et al. [Taylor and Letham (2018)]. Its software is available in Python and R for forecasting time series data. The prophet is based on a model in which non-linear weekly and annual seasonality are taken into account, as well as during holidays. Some of the strengths of the Prophet model are its strengths against lost data, large outliers, and the shifts in the trends. Besides, it can produce well enough estimate of the mixed data without spending manual effort. The prophet software has its special data structure in order to handle with the time series. To create estimates, it needs two main columns called "ds" and "y". "ds" is the actual times of the time series and "y" is the corresponding values. It predicts two main things i)  $\hat{y}$ , estimates of the model ii) the lower limit of  $\hat{y}$ , and iii) the upper limit of  $\hat{y}$ .

### 2.4 Long-short term memory (LSTM)

Long Short-Term Memory (LSTM) networks are created based on an extension for recurrent neural networks (RNN). As different from the traditional neural network, LSTM is designed to take important things learned from experiences previously occurred into consideration. The more detailed mathematical foundation of the LSTM model can be found in Lipton et al. [Lipton, Berkowitz and Elkan (2015)]. However, the main formulation of the model is given in Eqs. (12)-(17).

$$\hat{c}^{<t>} = g(W_{ca}a^{<t-1>} + W_{cx}x^{<t>} + b_c) \quad (12)$$

$$\Gamma_u = g(W_{ua}a^{<t-1>} + W_{ux}x^{<t>} + b_u) \quad (13)$$

$$\Gamma_f = g(W_{fa}a^{<t-1>} + W_{fx}x^{<t>} + b_f) \quad (14)$$

$$\Gamma_o = g(W_{oa}a^{<t-1>} + W_{ox}x^{<t>} + b_o) \quad (15)$$

$$c^{<t>} = \Gamma_u \hat{c}^{<t>} + \Gamma_f c^{<t-1>} \quad (16)$$

$$a^{<t>} = \Gamma_o c^{<t>} \quad (17)$$

where  $\hat{c}^{<t>}$  refers to the value of the memory cell. In other words, it is the "important" information from the previous time step.  $W_{ca}$  is the weight parameters for the memory cell.  $\Gamma_u$ ,  $\Gamma_f$ , and  $\Gamma_o$  refers to the update, forget, and output gate and, respectively,  $W_{ua}$ ,

$W_{fa}$ , and  $W_{oa}$  weights parameter of them. If  $\Gamma_u$  takes the value of 1,  $\hat{c}^{<t>}$  is going to be updated. In another word, the new “important” information will be stored. Based on weight parameters,  $\Gamma_u$ ,  $\Gamma_f$ , and  $\Gamma_o$  will be recalculated and the output of a neuron can be calculated based on the Eqs. (13)-(15).

### 3 Computational results

In this section, we have used classical regression models and machine learning models to create a prediction model for the daily COVID-19 cases of the seven countries. In order to compare used methods, we have evaluated the model performances with the Root Mean Squared Error (RMSE) evaluation metrics. The data used in the study was obtained from the COVID19 website of John Hopkins University (<https://coronavirus.jhu.edu/map.html>).

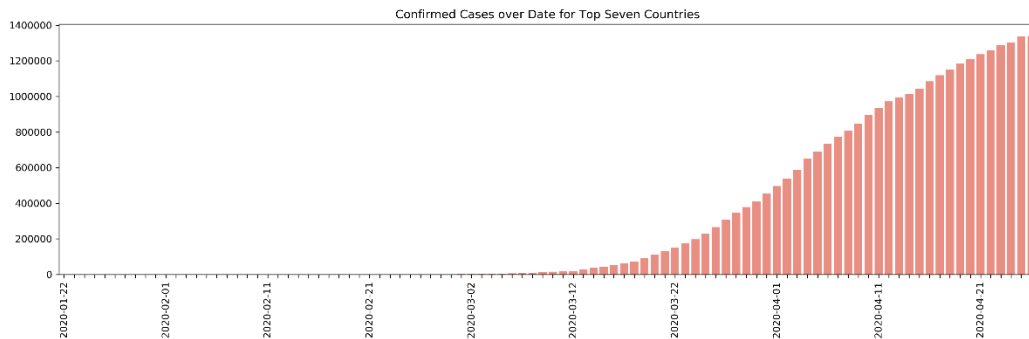
#### 3.1 COVID cases in the top seven countries

Through the study, the top seven countries in terms of the number of cases were considered. These countries are the USA, Spain, Italy, France, Germany, UK, and Turkey. As of April 28, 2020, there is a total of 2,919,40 confirmed cases worldwide, and 1974244 of them are coming from these seven countries. The following Tab. 1 shows the basic numbers regarding those seven countries.

**Table 1:** Clusters of the top seven countries

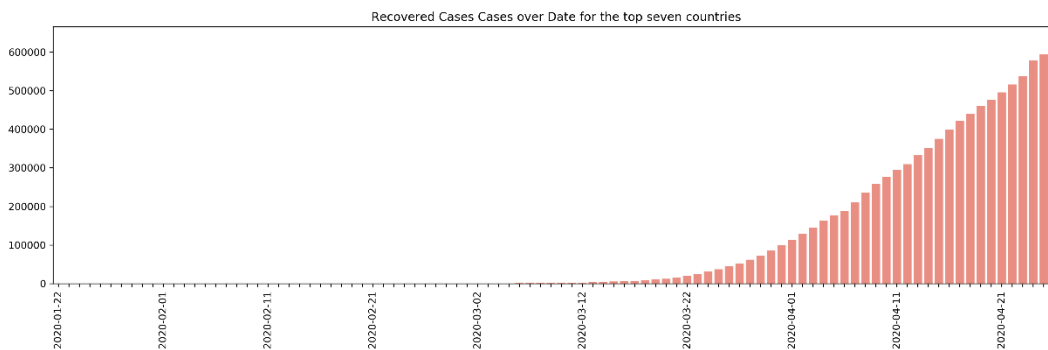
Cases	Total numbers
Total number of Confirmed Cases	2052183
Total number of Recovered Cases	513577
Total number of Deaths Cases	164281
Total number of Active Cases	1374325
Total number of Closed Cases	677858
The approximate number of Confirmed Cases per Day	20941
The approximate number of Recovered Cases per Day	5241
The approximate number of Death Cases per Day	1676

Following Fig. 1 illustrates the confirmed cases in all chosen seven countries. As shown, the total number of cases is increasing every day, but towards the end of April, the rate of increase is relatively decreasing. The historical flow of this situation is shown in Fig. 1, from the day of the first appearance of the virus until April 28.



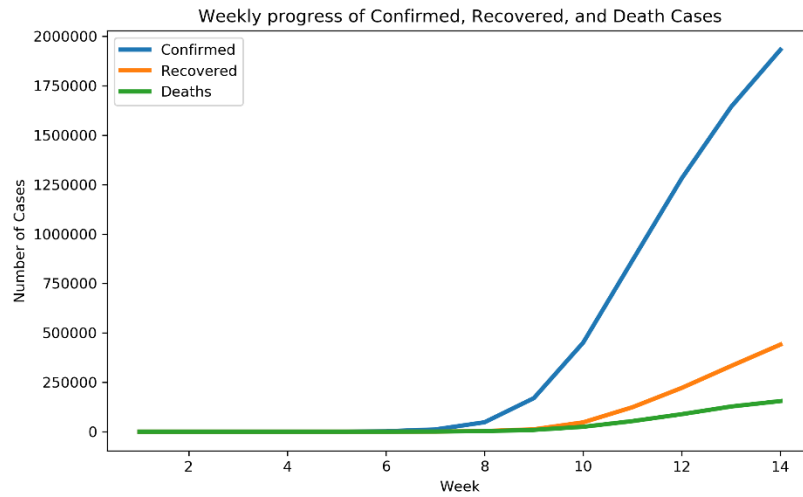
**Figure 1:** Confirmed Cases over date for the top seven countries

Also, in parallel with the total number of confirmed cases, the number of recovered patients are increasing more and more. The historical flow of this situation is shown in Fig. 2, from the day of the first appearance of the virus until April 28.



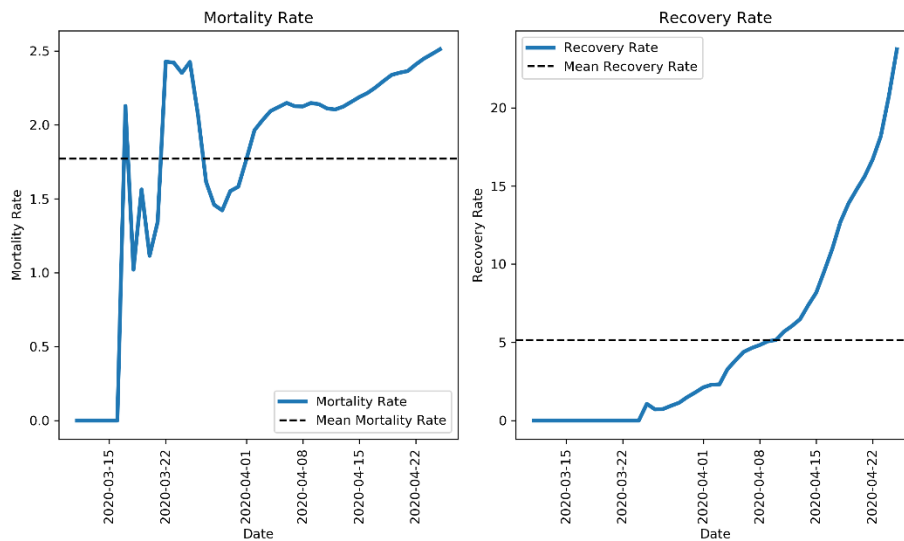
**Figure 2:** Recovered Cases over date for the top seven countries

And, Fig. 3 shows the weekly progress of confirmed, recovered, and death cases for the top-seven countries. The cases where the rate of increase is minimal concerning this figure are cases where the virus leads to death. Cases resulting in death are followed by the recovered and total number of cases, respectively. This may be due to the fact that the virus revealed in statistics has a more lethal effect on those over 65 years of age (i.e., according to Verity et al. [Verity, Okell, Dorigatti et al. (2020)], about 81% of the patients who died are over 60 years old). In other words, a small part of the total population is affected much more. Therefore, increases in mortality rate are expected to be less than in total and recovered cases.



**Figure 3:** Weekly progress of confirmed, recovered, and death cases

Finally, the mortality rate and recovered rate of the COVID-19 virus in those seven countries is shown in the following Fig. 4. The mortality rate is calculated by dividing the total number of deaths to the total number of confirmed cases and the recovery rate is calculated by dividing the total number of recovered cases to the total number of confirmed cases. As can be seen from Fig. 4, the recovery rate is progressing much faster than the mortality rate and is much higher than average from the beginning of April. This is the main source of optimistic scenarios for the future.

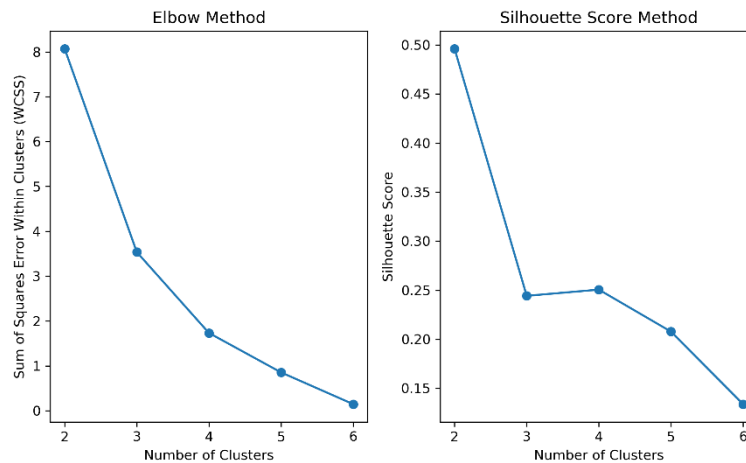


**Figure 4:** Average mortality and recovery rate over time



### 3.2 Clustering countries

To cluster seven countries, we have first tried to find the optimum number of clusters. To do this, we have used the elbow method and the silhouette score. While doing this, confirmed, recovered, and death cases are used as the features of the data. As shown in Fig. 5, both methods choose the optimum number of clusters as 3.



**Figure 5:** The results of Elbow and Silhouette methods

By setting the number of clusters as 3, we have applied the k-means clustering algorithm and results are shown in Tab. 2. Based on Tab. 2, Spain, Italy, and France in one cluster, Germany, UK, and Turkey in one cluster, and the US is in another cluster.

**Table 2:** Clusters of the top seven countries

Country	Confirmed	Recovered	Deaths	Mortality Rate(%)	Recovery Rate (%)	Clusters
US	938154	100372	53755	5.73	10.70	1
Spain	223759	95708	22902	10.23	42.77	2
Italy	195351	63120	26384	13.50	32.31	2
Germany	161644	45372	22648	14.01	28.07	0
France	156513	109800	5877	3.75	70.15	2
UK	149569	774	20381	13.62	0.52	0
Turkey	107773	25582	2706	2.51	23.73	0

### 3.3 Prediction of confirmed cases and deaths

In this section, we have created multiple prediction models for confirmed cases and deaths for the top seven countries. The following Tab. 3 gives the RMSE scores of SVM, Holt's Winter, Facebook's Prophet, and LSTM methods for those seven countries. To train the models we have used 95% of the dataset as the training set and 5% of it is used for testing.

**Table 3:** RMSE scores of all methods for each country

Method	Country	RMSE (Confirmed)	RMSE (Deaths)
SVM	US	83024.20	7766.17
	Spain	97678.44	10524.39
	Italy	116126.89	15366.96
	Germany	90907.79	387.40
	France	58132.31	4288.96
	UK	21895.26	1982.66
	Turkey	61791.23	1301.58
	Holt-Winters	Country	RMSE (Confirmed)
US		2565153.55	5545.41
Spain		1936.30	1757.06
Italy		10256.48	514.49
Germany		20932.14	2366.92
France		31714.37	3048.69
UK		9508.24	2004.49
Turkey		2364.57	115.44
Prophet	Country	RMSE (Confirmed)	RMSE (Deaths)
	US	<b>1783.23</b>	<b>535.70</b>
	Spain	<b>633.041</b>	<b>116.79</b>
	Italy	<b>704.71</b>	<b>74.58</b>
	Germany	<b>1041.48</b>	<b>44.54</b>
	France	<b>6565.46</b>	<b>371.72</b>
	UK	<b>528.46</b>	<b>103.23</b>
	Turkey	<b>463.90</b>	<b>4.79</b>
LSTM	Country	RMSE (Confirmed)	RMSE (Deaths)
	US	31888.16	1925.50
	Spain	7254.94	712.15
	Italy	1013.52	288.12
	Germany	5447.82	386.29
	France	6996.44	595.74
	UK	6054.15	217.42
	Turkey	6239.93	222.46

Based on Tab. 3, the best methods are the Prophet forecasting model with the lowest RMSE score for all countries. In addition to this, we have used the same methods for

future predictions. By using trained models, we have forecasted possible deaths until May 2, 2020. As of the preparation of this paper, we know the actual number of deaths until April 28. Thus, the actual number of columns have values until April 28.

**Table 4:** Future forecasting of number of deaths for Turkey

Dates	Actual	SVM	Holt's Winter	Prophet	LSTM
26.04.2020	2805	4804.249	2694.299	2852.495	2040.48
27.04.2020	2900	5341.946	2814.558	2970.77	2057.75
28.04.2020	2992	5927.405	2835.609	3091.71	2068.79
29.04.2020		6563.739	2960.864	3207.084	2076.91
30.04.2020		7254.193	3079.83	3328.117	2083.02
1.05.2020		8002.149	3211.313	3446.909	2086.49
2.05.2020		8811.127	3334.687	3565.558	2089.54

**Table 5:** Future forecasting of number of deaths for US

Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	54881	48772.57	46204.26	55963.68	47550
27.04.2020	56259	51412.08	48345.54	58137.78	48013
28.04.2020	58355	54163.89	51316.38	60428.06	48237
29.04.2020		57031.54	53383.97	62727.99	48417
30.04.2020		60018.66	54142.92	65262.33	48555
1.05.2020		63128.95	55084.23	67676.43	48648
2.05.2020		66366.16	57293.12	69938.06	48867

**Table 6:** Future forecasting of number of deaths for Spain

Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	23190	38670.38	26352.55	23735.62	20938
27.04.2020	23521	41000.57	27292.91	24212.44	21071.38
28.04.2020	23822	43441.69	27975.4	24637.88	21182.32
29.04.2020		45997.67	28605.97	25128.89	21322.35
30.04.2020		48672.5	29122.77	25639.49	21431.77
1.05.2020		51470.3	29611.7	26142.89	21563.53
2.05.2020		54395.24	30586.88	26570.12	21717.98

**Table 7:** Future forecasting of number of deaths for Italy

Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	26644	48381.82	27928.93	27212.48	24841.58
27.04.2020	26977	51261	28606.4	27720.48	25002.49
28.04.2020	27359	54275.64	29158.21	28251.36	25146.26
29.04.2020		57430.47	29717.14	28772.17	25278.67

30.04.2020	60730.32	30236.56	29267.09	25399.10
1.05.2020	64180.13	30772.72	29820.9	25513.91
2.05.2020	67784.96	31461.03	30345.77	25620.63

**Table 8:** Future forecasting of number of deaths for Germany

Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	5976	6925.481	2243.168	6125.034	4778.20
27.04.2020	6126	7325.445	5154.602	6346.662	4812.61
28.04.2020	6314	7743.587	2969.115	6555.685	4858.68
29.04.2020		8180.517	1429.647	6808.29	4906.05
30.04.2020		8636.863	2110.675	7036.501	4927.56
1.05.2020		9113.265	2078.928	7253.096	4944.03
2.05.2020		9610.377	4771.584	7430.779	4938.09

**Table 9:** Future forecasting of number of deaths for France

Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	22890	30414.29	28359.11	24922.14	20665.39
27.04.2020	23327	32093.06	30736.68	25650.31	20844.80
28.04.2020	23694	33844.8	30986.5	26444.25	20992.87
29.04.2020		35671.87	32647.12	27223.79	21105.86
30.04.2020		37576.68	32983.99	28086.71	21208.70
1.05.2020		39561.67	33190.68	28888.01	21273
2.05.2020		41629.36	35800.76	29642.24	21319.40

**Table 10:** Future forecasting of number of deaths for UK

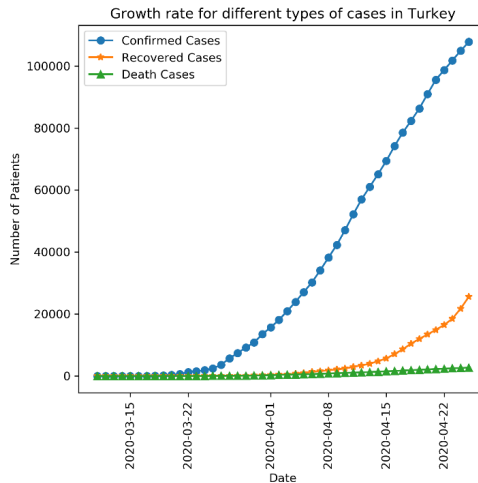
Dates	Actual	SVM	Holt-Winters	Prophet	LSTM
26.04.2020	20794	24674.15	24428.02	21376.81	21577.93
27.04.2020	21157	26150.07	25012.94	22094.2	21759.08
28.04.2020	21745	27695.43	26643.13	22869.08	21901.23
29.04.2020		29312.65	27647.57	23661.7	22010.96
30.04.2020		31004.22	28679.87	24451.13	22093.37
1.05.2020		32772.65	29295.29	25256.04	22142.46
2.05.2020		34620.56	29805.77	26064.79	22169.23

As shown in Tabs. 4-10, the Prophet forecasting model estimates the closest values to the actual values.

### **3.4 Deep analysis of Turkey**

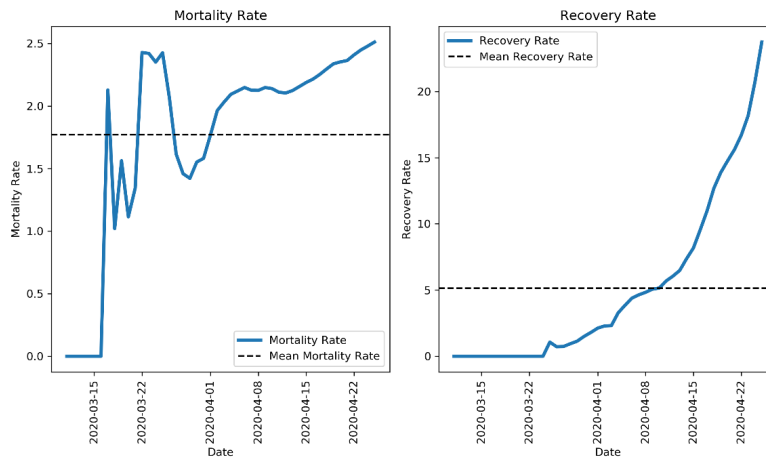
Turkey is one of the seven countries most affected by the virus epidemic. Compared to other countries, it managed to prevent the spread of the virus much faster by taking nationwide measures against the virus earlier. In this part, Turkey's virus data are examined in more detail and a projection for the future is revealed. As of April 26, 2020,

there are 110130 confirmed cases, 29140 recovered patients, 2805 deaths nationwide. Fig. 6 shows the growth rate for different types of cases in Turkey. One of the most remarkable points is possible to see the partial effect of curfew implemented in Turkey as of April 22 from the change of confirmed cases during the last 4 days. This suggests that social seizure is one of the most important factors in preventing the spread of the virus.



**Figure 6:** Growth rate for different types of cases in Turkey

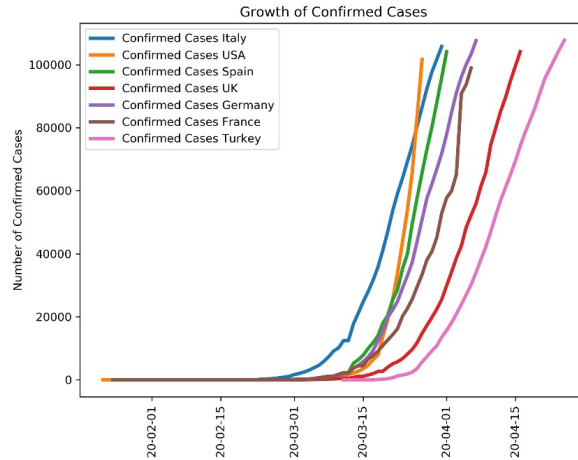
On the other hand, Fig. 7 shows the rate of death and recovery from the date of the first death in Turkey. It suggests that by the end of April, the recovery rate increased much faster than the mortality rate.



**Figure 7:** The growth rate of mortality and recovery for different types of cases in Turkey

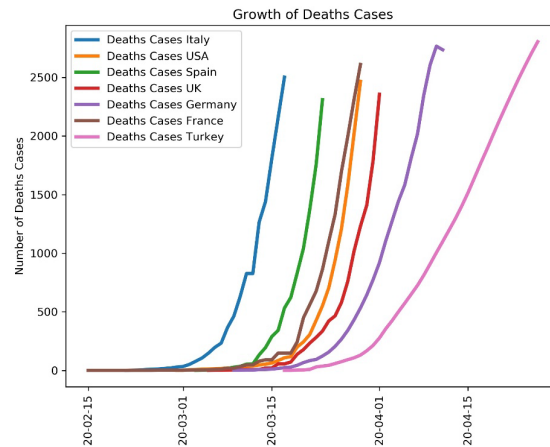
As shown in Tab. 2, Turkey is the country with the lowest mortality rate of 2% among seven countries. It is also ranked three by 23.73% compared to the average recovery rate. Besides, when we compare the total number of cases in other countries, as shown in Fig. 8, the date of occurrence of the first case in Turkey is much later than the other six

countries. It took 61, 66, 61, 77, 71, and 74 days in Italy, USA, Spain, UK, Germany, and France to reach confirmed cases equivalent to Turkey. One of the points to be considered here is that the number of cases has increased rapidly in a very short time after the first case in Turkey. One of the reasons may be that Turkey's daily tests have reached about forty thousand in a very short time (<https://covid19.saglik.gov.tr/>).



**Figure 8:** Comparison of Turkey with the other six countries in terms of confirmed cases

Another comparison is the comparison between Turkey and six other countries in terms of number of deaths. As shown in Fig. 9, Turkey appears to be one step ahead of other countries in this regard. The total number of mortal cases in Turkey is 2805 from the date of occurrence of the first mortal case in Turkey until April 26. And these deaths took place in a total of 41 days. It took 26, 30, 21, 28, 34, and 44 days in Italy, USA, Spain, UK, Germany, and France to reach confirmed cases equivalent to Turkey. One of the points to be considered here is that the number of cases has increased rapidly in a very short time after the occurrence of the first case in Turkey. As in confirmed cases, it may be the result of that the daily test numbers in Turkey have reached about forty thousand in a very short time.



**Figure 9:** Comparison of Turkey with the other six countries in terms of deaths

Measures against the coronavirus in Turkey have been firmly implemented by the government since mid-March and are still underway. Recently, however, a little loosening of these prohibitions has begun to be discussed in the national press by creating a prediction model for the future by the Prophet forecasting model, which is the lowest margin of error among the methods applied, we would like to see how applicable and how risky those prospective decisions are and aimed to predict what period the mortality rate would go toward zero. Fig. 10 below shows the 150-days forecasting model for Turkey. Based on Fig. 10, when considering the lower limit of the model, it is projected that the number of deaths in Turkey will approach zero from mid-September to early October. Besides, Fig. 10 shows the prediction of the number of cases for future 150-days. Based on Fig. 10, when considering the lower limit of the model, it is projected that the total number of confirmed cases in Turkey will approach zero from the beginning of August.

Obviously, future predictions are based on predicted values after some certain point. That's why the error rate might be higher than real values. Thus, the estimates of the short term, such as weekly, might give better guidance to policymakers. Based on those estimates, new measures can be implemented, or existing measures can be revised to minimize the effect of the COVID-19.

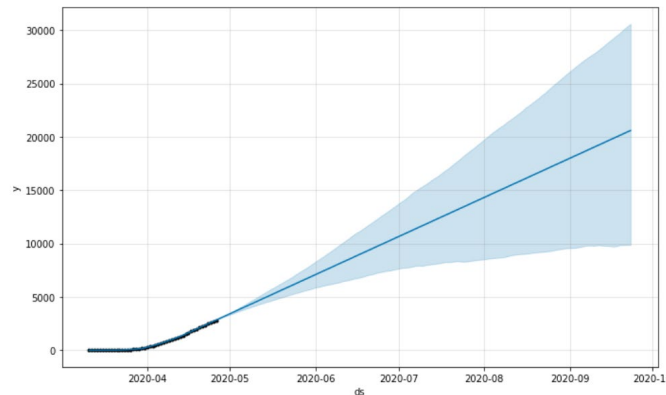


Figure 10: 150-days forecasting model for death cases in Turkey

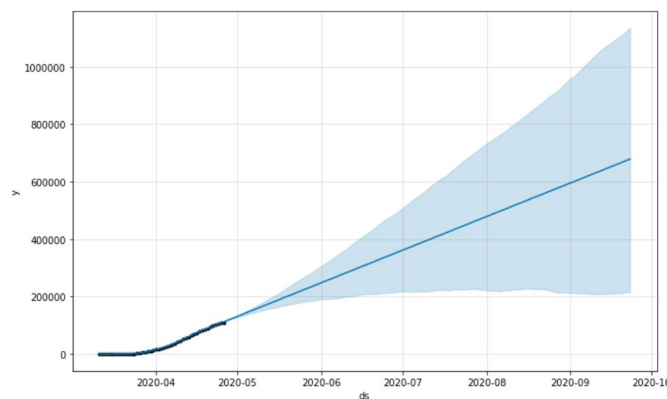


Figure 11: 150-days forecasting model for total confirmed cases in Turkey

#### **4 Conclusions and discussions**

In this study, we have analyzed the COVID-19 data of top-seven countries in terms of confirmed cases. We have first analyzed all seven countries in terms of basic descriptive statistics. Then we have clustered those seven countries. Based on the elbow method and silhouette scores US differs from the other six countries while Spain, Italy, and France in one cluster and UK, Germany, and Turkey in another cluster.

Then, we have used different prediction models for those top-seven countries. These models are Support Vector Machines (SVM), Holt-Winters, Facebook's Prophet, and Long-Short Term Memory (LSTM) Among those models, the Facebook's Prophet forecasting model gives the lowest RMSE score for all countries. Besides, by using the Facebook's Prophet method, we estimated Turkey's next 150 days of deaths and confirmed cases. Based on these predictions, it is projected that the number of deaths in Turkey will approach zero from mid-September to early October and it is projected that the total number of confirmed cases in Turkey will approach zero from the beginning of August. In the light of these predictions, loosening the measures taken to minimize the effects of coronavirus epidemic in Turkey in a short period might cause the second wave of epidemic in Turkey.

In future studies, using these forecasting models, a worldwide planning can be planned to decide how resources can be distributed across countries. For example, health workers or health materials such as respirators in a country where the epidemic is predicted to end early, can be transferred to a country where the epidemic is expected to end later. By doing this, the effects of the virus can be alleviated or finished all over the world.

**Funding Statement:** The author(s) received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### **References**

- Allam, Z.; Jones, D. S.** (2020): On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. *Healthcare*, vol. 8, no. 1, pp. 46.
- Barstugan, M.; Ozkaya, U.; Ozturk, S.** (2020): *Coronavirus (COVID-19) Classification Using CT Images by Machine Learning Methods*. <https://arxiv.org/abs/2003.09424>.
- Chauhan, V. K.; Dahiya, K; Sharma, A.** (2019): Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review*, vol. 52, no. 2, pp. 803-855.
- Deb, S.; Majumdar, M.** (2020): *A Time Series Method to Analyze Incidence Pattern and Estimate Reproduction Number of COVID-19*. <https://arxiv.org/abs/2003.10655>.
- Funk, S.; Camacho, A.; Kucharski, A. J.; Eggo, R. M.; Edmunds, W. J.** (2018): Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, vol. 22, pp. 56-61.



**Hansun, S.; Charles, V.; Indrati, C. R. (2019):** Revisiting the Holt-Winters' additive method for better forecasting. *International Journal of Enterprise Information Systems*, vol. 15, no. 2, pp. 43-57. <https://arxiv.org/abs/2002.12298>.

<https://www.medrxiv.org/content/10.1101/2020.03.27.20045237v1>.

**Jenny, P.; Jenny, D. F.; Gorji, H.; Arnoldini, M.; Hardt, W. D. (2020):** *Dynamic Modeling to Identify Mitigation Strategies for COVID-19 Pandemic*.

**Jiang, X.; Coffee, M.; Bari, A.; Wang, J.; Jiang, X. et al. (2020):** Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, vol. 63, no. 1, pp. 537-551.

**Li, L.; Qin, L.; Xu, Z.; Yin, Y.; Wang, X. et al. (2020):** Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest ct. *Radiology*, 200905.

**Lipton, Z. C.; Berkowitz, J.; Elkan, C. (2015):** *A Critical Review of Recurrent Neural Networks for Sequence Learning*. <https://arxiv.org/abs/1506.00019>.

**Liu, Z.; Magal, P.; Seydi, O.; Webb, G. (2020):** *Predicting the Cumulative Number of Cases for the COVID-19 Epidemic in China from Early Data*.

**Pirouz, B.; Shaffiee Haghshenas, S.; Shaffiee Haghshenas, S.; Piro, P. (2020):** Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability*, vol. 12, no. 6, pp. 2427.

**Randhawa, G. S.; Soltysiak, M. P.; El Roz, H.; De Souza, C. P.; Hill, K. A. et al. (2020):** Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS One*, vol. 15, no. 4, e0232391.

**Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R. et al. (2020):** Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, vol. 5, pp. 256-263.

**Taylor, S. J.; Letham, B. (2018):** Forecasting at scale. *The American Statistician*, vol. 72, no. 1, pp. 37-45.

**Verity, R.; Okell, L. C.; Dorigatti, I.; Winskill, P.; Whittaker, C. et al. (2020):** Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).

**Wu, J. T.; Leung, K.; Leung, G. M. (2020):** Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, vol. 395, no. 10225, pp. 689-697.