American Interdisciplinary Journal of Business and Economics

ISSN: 2837-1909 Impact Factor : 4.6 Volume. 9, Number 4; October-December, 2022; Published By: Scientific and Academic Development Institute (SADI) 8933 Willis Ave Los Angeles, California https://sadipub.com/Journals/index.php/aijbe



PROPHET MODEL'S EFFICIENCY IN SHORT-TERM COVID-19 CUMULATIVE CASE PROJECTIONS: G7 COUNTRIES

Yeliz Zoğal

Ankara Haci Bayram Veli University

Abstract: The Covid-19 pandemic has had a significant impact on the health and well-being of people across the globe, as well as the global economy at large. It has become essential to predict the spread of infectious diseases like Covid-19 to understand its impact on public health and the economy. This study analyzes short-term predictions of Covid-19 cases in G7 countries using the Prophet model. The model uses a trend function, a seasonality function, and a holiday function to generate accurate short-term predictions. The study compares the predictions for G7 countries and finds that Canada and Germany had the lowest root mean square error (RMSE) values. The economic and financial impacts of the pandemic on global supply chains, job losses, and business closures are also analyzed. The study highlights the significant increase in public debt due to large-scale fiscal stimulus packages implemented by governments to mitigate the economic impact of the pandemic. The study emphasizes that accurate predictions of the spread mechanism are crucial for managing the pandemic effectively and mitigating its impact. The literature review of various models indicates the importance of accurate predictions and the difficulties in creating them. The study recommends the use of machine learning models like the Prophet model to generate accurate short-term predictions to combat the Covid-19 pandemic's spread.

Keywords: Covid-19, Prophet model, short-term predictions, G7 countries, root mean square error, economic impacts, global supply chains, fiscal stimulus packages, machine learning models.

1. Introduction

In December 2019, it was reported to the world that a virus associated with severe acute respiratory syndrome began to spread across China's Wuhan city. This virus, later named Covid-19, has started to spread worldwide, and cases have become unpreventable. In the WHO Director General's statement on COVID-19 dated March 11, 2020, he stated that in two weeks, the number of Covid-19 cases increased thirteen times, the number of affected countries tripled, there were more than 118,000 cases in 114 countries and 4,291 people died.

The unpreparedness of countries for infectious diseases has caused them to struggle with a lack of capacity, resources, and determination in general, especially in the health sector. In addition to the severe impact of COVID-19 on healthcare systems, the pandemic has had a significant and rapidly escalating impact on the world economy and businesses.

The COVID-19 pandemic has been ongoing for more than three years and continues to cause significant health and economic losses. According to official estimates, more than 6 million people have died from the virus, with studies estimating the actual death toll to be much higher, ranging from 16 to 20 million, which is approximately equal to that of World War I. According to the IMF's World Economic Outlook (2022), the cumulative output loss from the pandemic through 2024 is projected to be about \$13.8 trillion and it is likely that the actual loss will be even higher. According to the World Economic Outlook (WEO) report published in October 2021, the economic contraction for G7 countries is as follows: Germany at 4.6%, France at 8%, Italy at 8.9%, United Kingdom at 9.8%, United States at 4.3%, Canada at 5.3%, and Japan at 4.6%. In G7

countries, the pandemic has led to a significant increase in unemployment and a decrease in economic activity. G7 countries have also seen a decline in consumer spending, particularly in sectors such as travel and tourism. The economic impact of the COVID-19 pandemic has been very severe and has occurred much more rapidly than the 2008 global financial crisis (GFC) and the Great Depression. The rapid

spread of the virus and the measures put in place to control it have led to widespread job losses and business closures, disruptions in global supply chains, and a decline in economic activity. In the 2008 GFC and the Great Depression, stock markets collapsed by 50% or more, credit markets froze up, massive bankruptcies followed, unemployment rates soared above 10%, and GDP contracted at an annualized rate of 10% or more, but all of this took around three years to play out. While in the current crisis, similarly dire macroeconomic and financial outcomes have materialized in a much shorter period, in some cases, just three weeks. The speed of the economic downturn caused by the COVID-19 pandemic is largely due to the rapid spread of the virus and the measures put in place to control it. The pandemic has also led to a significant increase in public debt as governments have implemented large-scale fiscal stimulus packages to mitigate the economic impact of the pandemic.

The COVID-19 crisis has brought an unprecedented shock to the labor market and has led to an unemployment crisis. Restrictions and lockdowns implemented since March 2020, as well as the decline in demand caused by the pandemic, have led to millions of job losses around the world. Although some measures taken to prevent the spread of the virus have allowed some people to work from home, in most countries it has caused unemployment. The visible contraction in the economy is reflected in the unemployment rates. According to the IMF's report in 2021, the unemployment rate in G7 countries is Germany at 3.8%, France at 8%, Italy at 9.3%, United Kingdom at 4.5%, United States at 8.1%, Canada at 9.6% and Japan at 2.8%. In addition to the hike in unemployment rates, the profound effect has started to increase income inequalities and poverty, which was estimated as more 71 million people as of March 2021.

The outbreak of COVID-19 has had a significant impact on global supply chains. The disruption of transportation and production caused by lockdowns and other measures taken to contain the spread of the virus has led to shortages of goods and materials, delays in delivery times, and increased costs for businesses. Consumers have also been affected by the disruption of supply chains, with many facing shortages of essential goods and higher prices for products. The domino effect of broken supply chains is that it can cause a ripple effect throughout the economy, affecting not just producers and consumers but also other businesses that rely on them. This can lead to job losses, reduced economic activity, and other negative consequences.

On the supply side, production chains have been disrupted, while on the demand side, consumption and investment spending have been negatively affected. This is likely to exacerbate the ongoing economic downturn, making it more pronounced. To address both the pandemic and economic downturn, governments have been urged to "Go big. Act fast. Keep the lights on" by economist Richard Baldwin, who argues that combining restrictive policies that reduce production with stimulus policies that maintain spending will create supply-side problems and lead to cost-driven inflation [2]. In other words, the idea that the global downturn can be revived only by increasing credit and borrowing more heavily for consumption is an illusion.

The COVID-19 pandemic has had a significant impact on healthcare economics around the world. In the short run, healthcare facilities have been overwhelmed by the influx of patients, leading to increased costs for inpatient and outpatient care. This has been compounded by the need for additional resources such as personal protective equipment and additional staff to handle the increased workload. In the long run, the economic impact of the pandemic on healthcare systems may be even more severe. The prolonged disruption of healthcare services and the increased demand for care could lead to higher costs for both patients and healthcare providers, as well as longer wait times for appointments and procedures. Additionally, the pandemic has led to a decline in revenue for many healthcare providers and hospitals, which could lead to financial difficulties and closures. This could lead to further strain on the healthcare system in the long run as the population increases and aging.

The listed reasons above make understanding the spread mechanism and forecasting essential for effectively managing the pandemic and minimizing its impact on public health and the economy.

In the literature, many studies deal with understanding the spread mechanism. For instance, the mathematical SIR (Susceptible, Infected and Recovered) model provides differential solutions by dividing the total

population into different groups. The model describes the flow of individuals between these compartments based on certain assumptions, such as the rate at which infected individuals infect susceptible individuals (the infection rate) and the rate at which infected individuals recover or are removed from the population (the removal rate). The SIR model can be used to estimate the number of individuals who will be infected and recovered over time, as well as the peak of the epidemic. Although the characteristics of the infectious disease shape the model, it does not provide satisfactory results in the early stages of the disease. The econometric time series model, ARIMA (autoregressive integrated moving average), provides more realistic results as the data increase since it predicts the future of the variable with past values. The ARIMA model can be used to model and forecast time series data with a trend and/or seasonality. The ARIMA model is widely used in various fields such as finance, economics, engineering, and infectious diseases, and it is considered as one of the most powerful tool for time series forecasting. In addition, machine learning models (supervised learning, reinforcement learning models, deep learning models, ensemble models) are frequently used to predict infectious diseases. The present study aims to provide a five-day Covid-19 prediction for G7 (Canada, France, Germany, Italy, Japan, Japan, UK and USA) countries with the Prophet model.

Moreover, it allows comparisons with the RMSE (root mean square error) statistic to evaluate the performance of the analysis results. G7 countries were chosen because data sharing problems are less for these countries, and doubts about the accuracy of the number of cases announced are minimal. Based on the results obtained with the Prophet model, the results closest to reality with the lowest RMSE value were obtained first for Canada and then for Germany. Likewise, it was found that the prediction values for the first day generally have lower RMSE. In other words, RMSEs increase as we move away from the actual data for prediction.

The next section of the study presents the literature review and the theoretical framework for the model. In the third section, empirical findings will be presented. The last section of the study will provide a general evaluation of the research and analysis.

2. Literature Review and Theoretical Background

2.1. Literature Review

It is explained the results of SIR, SEIR, SEIRU, SIRD, SLIAR, ARIMA, ARIMA and SIDARTHE models used in the prediction of the spread mechanism, peak and decline of Covid-19 cases and the difficulties in creating predictions. The results of studies conducted with these models for different countries are graphed, and the performance of the models are compared with actual values and deviation value of predictions. Regarding this issue, it was shown that the highest deviation was found in the simple mathematical model and SIRD model for California.

The study used ARIMA model to forecast the trend of the COVID-19 outbreak in Italy, Spain, and France, which are the countries that were most affected by the pandemic in Europe, using data from the period of February 21 to April 15, 2020. The different past period ARIMA models were compared with the MAPE performance value and the ARIMA (0,2,1), ARIMA (1,2,0), and ARIMA (0,2,1) models were selected for the countries, respectively. With these selected models, short-term predictions were made for the period of April 16 to April 25, 2020. The study is showing that the ARIMA models can be used to effectively predict the trend of the COVID-19 outbreak, which can help governments and healthcare providers to prepare better and allocate resources more efficiently.

It is aimed to obtain forecasts for two days, namely February 11 and 12, using an ARIMA model with the Covid-19 case counts for the period between January 20 and February 10, 2020, published by Johns Hopkins University. They emphasized that case definition and data collection for cases should be simultaneous to obtain more realistic predictions.

It is aimed to predict the number of positive cases of different influenza (for H1N1 and H3N2 viruses) that may occur in 2016 with the number of pediatric cases of the influenza season between 2007 and 2015. To this end, they used both ARIMA and seasonal ARIMA, that is, SARIMA (an ARIMA model that can capture seasonal effects-seasonal autoregressive integrated moving average). The prediction results of the models are evaluated according to performance criteria. Accordingly, it is shown that the ARIMA model gives better and more realistic results than the SARIMA model in case prediction.

Time series models are used (ARIMA and SARIMA) and a machine learning model (Prophet model). The prediction is based on daily and cumulative Covid-19 data for the United States, India, and Brazil, and they

obtain short-term prediction results. Specifically, the Prophet model, which can capture periodic features in the data, gives better results for the US forecasts, while the ARIMA model gives better results for Brazil and India, whose cumulative cases tend to grow. The SARIMA model also captures daily cases' seasonal characteristics and provides better prediction results

It is used ARIMA from time series models and Prophet, GLMNet (generalized linear model elastic net), random forest and XGBoost from machine learning models to predict Covid-19 cases. Based on the results of the analysis, ARIMA and Prophet models are more appropriate and ideal forecasts for the countries in question, especially the ARIMA model gives better results in Afghanistan, Bangladesh, India, Maldives, and Sri Lanka. They explained that the random forest machine learning model was excluded due to its poor fit to the data set

It is obtained short-term prediction results with Covid-19 data for India for the period January 30 - December 7, 2020. The results were obtained using ARIMA and some machine learning models such as Prophet, LSTM (long short-term memory), RNN (recurrent neural network), GRU

(Gated recurrent unit) and LSTM-GRU models. R2 and RMSE values were obtained for numerical comparison of the performance of the models. In conclusion, it is shown that the LSTM-GRU model is superior to the others with high R2 and low RMSE values [15].

It is used Covid-19 data for twenty countries and obtained short and long-run predictions using SEIR, polynomial regression, ARIMA and Prophet model. According to the prediction results, the polynomial regression model gives the best short-term predictions, while the SEIR model gives the best long-term predictions

It is developed a new exponential growth model in their study, arguing that improving epidemic models can be more helpful in explaining different periods of an epidemic. The model aims to characterize the stage of the epidemic, especially when it shows an upward trend, and to capture the changing epidemic profile for that period. To this end, the model is applied to eight different infectious diseases with various transmission routes in twenty different geographies and the same infectious disease (Ebola) in different periods. The results show that the growth rate of the same infectious disease changes over time and how different geographical and social conditions affect the growth rate. It is explained that the epidemic growth rate is primarily influenced by limited population contact structure, behavioural change over time or early control interventions.

In a study, the authors have highlighted the failures of models used to predict the spread of infectious diseases. The failure of models used to predict the spread of COVID-19, has made this situation even more pronounced. The reasons for this failure are poor data input, poor modeling, inaccurate and inconsistent assumptions, the predictors being overly sensitive, the distinctive features of the outbreak not yet fully determined and included in the models, the lack of accuracy of existing prevention measures, lack of transparency of data, lack of determining parameters, and reporting errors. However, solutions for some of these issues have been proposed such as making wave predictions instead of point predictions and selecting models that are developed and expanded based on performance results

In a study, the authors aimed to use various time series forecasting models such as Prophet, Holt-Winters, LSTM, ARIMA, and ARIMA-NARNN to predict short-term daily and cumulative case forecasts of Covid-19, model the general trend of the outbreak, and model the time series based on linear and non-linear features. The results obtained were compared with various statistical measurements. In this regard, it was reported that the models showed good performance, but the ARIMA and NARNN (ARIMA-NARNN) hybrid combination had the best performance

In a study, the aim is to estimate the extent of the COVID-19 outbreak in Pakistan and case forecast predictions using ARIMA, Diffusion, SIRD and Prophet Models. The short-term forecast results obtained show similarities and indicated that the highest number of infectious cases could be reached between June 2020 and July 2020. Due to this reason, it is conveyed that most of the population is under the threat of COVID-19 and that the measures taken by the government should be reviewed and improved

2.2. Theoretical Background

Prophet model is a time series forecasting model developed by Facebook's Core Data Science team in 2018. It is designed to make forecasting future data points as simple as possible and is particularly well-suited for business time series data. Prophet is a procedure for forecasting time series data based on an additive model

where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects [13]. The model has three different components

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$
(1)

Here, y(t) is the number of cases, g(t) is the trend, that is, the trend function that captures non-periodic changes in the time series, s(t) is the seasonality, that is, the part of the time series that captures periodic (weekly or yearly) changes, h(t) is the part that represents holidays that occur at irregular intervals, and ε_t is the error term that includes specific changes that cannot be covered or handled by the model [17]. The functions handled in the model are as follows:

$$g(t) = 1$$
_____+ $e x p(-C k(t-m))$ (2)

As the growth in Facebook is similar to growth in a natural ecosystem (just like the increase in cases), a logistic growth model was added to capture the increase in the trend. C denotes the carrying capacity, k denotes the growth rate, and m denotes the offset parameter.

$$n\pi t \qquad s \qquad (3) \qquad (t) = \sum_{n=1}^{\infty} a_n \cos^{\frac{2n\pi t}{t}} + b_n \sin^{\frac{2}{t}} \qquad p \qquad p$$

Time series are often susceptible to multi-period seasonality due to the human behaviour they represent. This can be a five-day working week, vacation schedules or school holidays that follow each year [17]. To capture these periodic effects, the Fourier model is used. Here p is the periodic term the time series is expected to have regularly.

$$h(t) = Z(t)\kappa \tag{4}$$

$$Z(t) = [i(t \in D_i), \dots, i(t \in D_i)]$$
(5)

New year holidays, religious holidays or events that may have a country-wide impact are predictable shocks for time series models. For this reason, the impact of such holidays on the model is included daily. D_i is defined for each holiday *i* in the model. Therefore, that time series interval represents the presence of that holiday if it coincides with that period. The κ parameter is defined as the corresponding change parameter in the forecast if the day of that holiday changes in the time series.

Since the Prophet model includes both logistic and Fourier, it captures periodic waves in the observations more efficiently, allowing for relatively better predictions with outliers in the data. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well [13].

One of the frequently used statistical values for comparing the performance of model results is the root mean square error (RMSE). Here, y° is the prediction value, y is the actual value, and n is the total number of observations. RMSE is calculated as the square root of the mean of the squared differences between predicted values and actual values. The formula is as follows:

$$\sqrt{\sum (y^{-}y)^2} \tag{6}$$

п

The RMSE value is expressed in the same units as the original data, so it can be directly interpreted in terms of the problem being solved. The smaller the RMSE value, the better the model is at predicting the actual values. However, it should be noted that comparing RMSE values between different datasets or problems can be misleading as the scale of the data and the specific problem objectives can be different.

3. Empirical Findings

In the study, the daily number of Covid-19 cumulative cases for G7 countries was obtained from Our World in Data for 180 days, starting with the day of the first occurrence. The date range for each country varies according to the day the case first started. In the prediction analyses for each country, the forecasted first day is the prediction for the 181st day, the forecasted second day is the prediction for the 182nd day, the forecasted third day is the prediction for the 183rd day, the forecasted fourth day is the prediction for the 184th day, and the forecasted fifth day is the prediction for the 185th day.

 Table 1. Forecast and actual values

1st day 2nd day 3rd day 4th day 5th day

				$y^{(\text{forecasted})}$	112,220	112,754	113,281	113,779	114,163
Canada	a y	(actual)	112,281	112,663	113,399	113,	836	114,172	RMSE
$y^{(for}$	ecasted)		-		4.58	6.78	8.77	4.23	0.69
France	y (actual)	217,5	517		222,122	224,749	227,016	228,708	232,624
	218,753	219,8	344 —		343.20	446.94	534.56	654.42	946.02
	219,928	219,9	932		204,765	205,044	205,280	205,604	205,939
RMSE	2								
y^(for	ecasted)								
Germa	iny y	(actual)	204,964	205,269	205,609	206	242	206,926	RMSE
y^(for	ecasted)		· _	-	14.83	16.77	24,52	47.55	73.57
Italy	y (actual)	246,7	776		246,860	247,480	248,205	249,144	250,159
2	247,158	247,5	537 —		6.25	24.00	49.81	97.79	155.71
	247,832	248,0)70		25,877	26,562	27,300	28,103	28,961
RMSE	l								
y^(for	ecasted)								
			Japan	RMSF					
			$y^{(forecasted)}$						
			UK	y (actual)	301,455	302,301	303,181	303,942	304,685
			_	• • •	14,71	18.63	14.39	1.09	6.97
				v (actual)	301,982 25,680	303,226 26 312	304,626 27 107	306,014 28 088	307,628 28 867
RMSE	ì			y (actual)	39 30	68.92	27,107 107 73	154.46	20,007
$v^{(for})$	ecasted)		_		3 834 404	3.897.260	3.958.010	4.017.141	4.072.962
5 (101)	coustour	I	USA	y (actual)	3,828,431 3,5	896,716 3	,964,163 4	4,031,940	4,106,884

RMSE 445.20 40.55 458.62 1,103.05 2,528.40

For G7 countries, the predictions of the number of cumulative cases for the 181st, 182nd, 183rd, 184th, and 185th day, the actual number of cumulative cases of these days, and RMSEs for each day are reported in Table 1. For France, Germany, Italy, and the UK, the RMSE increases with distance from the period used for the prediction between day one and day five, while for Canada, Japan, and the USA, the RMSE varies. When analyzed in detail, the estimated number of coincidences for Canada for the five days ranged from 112,200 to 114,163. The actual number of cases during this period ranges from 112,281 to 114,172. The lowest RMSE value for Canada was obtained on the fifth day. For France, predicted values ranged from 222,122 to 232,624, and actual values ranged from 217,517 to 219,932. The lowest RMSE value for France was obtained on the first day and increased until the fifth day. For Germany, the estimated number of cases ranged from 204,765 to 205,939, while the actual cases ranged from 204,964 to 206,926. The lowest RMSE for Germany was obtained on the first day. For Italy, the predicted cumulative cases were 246,860 on the first day and 250,159 on the last day (the fifth day). The cumulative number of cases for the five days varies between 246,776 and 248,070. The lowest RMSE for Italy was obtained on the first day. The estimated cumulative number of cases for Japan varies between 25,877 and 28,961. The cumulative number of cases ranges from 25,680 on the first day to 28,867 on the fifth day. The lowest RMSE for Japan was obtained on the fourth day. For the UK, the predictions for the number of cases ranged from 301,982 to 307,628 for the five days. The actual number of cases ranged from 301,455 to 304,685. The lowest RMSE for the UK is obtained for the first day of the prediction. Finally, the estimated cumulative data in the USA are 3,834,404 for the first day and 4,072,962 for the fifth day. The actual cumulative cases during this period vary between 3,708,557 and 3,708,557. The lowest RMSE for the USA was obtained on the second day.

Yeliz Zoğal (2022)

The lower and upper bound predictions of the estimated numbers of cumulative cases according to the prophet model are reported in Table 2. For Canada, except for the fifth day, the predicted number of cases on all other days fell between the lower and upper bounds. For France, the number of cases on the fourth and fifth days was below the lower predicted value. In Germany, the number of cumulative cases remained within the predicted range. The number of cumulative cases in Italy on the first day only, in Japan on the fourth day, and in the USA on the second day only was within the predicted range. For the UK, the number of cumulative cases on all days was below the lower bound of the predicted range.

 Table 2. Interval of forecasted value

$\begin{array}{c ccccc} 112,112 & 112,640 & 113,158 & 113,637 & 113,979 \\ 112,309 & 112,871 & 113,401 & 113,930 & 114,373 \\ 214,108 & 215,628 & 218,392 & 220,465 & 224,020 \\ 230,679 & 233,821 & 235,715 & 236,962 & 240,967 \\ 204,238 & 204,441 & 204,610 & 204,761 & 204,764 \\ 205,315 & 205,611 & 205,950 & 206,579 & 207,308 \\ 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ 25,823 & 26,507 & 27,241 & 28,037 & 28,893 \\ 25,931 & 26,621 & 27,363 & 28,164 & 29,032 \\ 25,823 & 26,507 & 27,241 & 28,037 & 28,893 \\ 25,931 & 26,621 & 27,363 & 28,164 & 29,032 \\ 301,638 & 302,846 & 304,237 & 305,568 & 307,094 \\ any & 302,349 & 303,605 & 305,058 & 306,505 & 308,225 \\ 3,831,657 & 3,894,363 & 3,954,755 & 4,013,203 & 4,067,776 \\ y^{1} o w e r \\ y^{u} p p e r \\ \hline Forecasted & y^{1} o w e r \\ y^{u} p p e r \\ \hline Forecasted & y^{1} o w e r \\ y^{u} p p e r \\ \hline Forecasted & y^{1} o w e r \\ y^{1} o w $		<i>y</i> ^ <i>u p p e i</i>	r	442 450	110 (07	440.000
Forecasted $y^{10} w er$ $y^{10} v p p er$ Forecasted $y^{10} w er$ $y^{10} w p p er$ Forecasted $y^{10} w er$ $y^{10} w p p er$ Forecasted $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ Forecasted $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ Forecasted $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ $y^{10} w er$ Forecasted $y^{10} w er$ $y^{10} w er$		112,112	112,640	113,158	113,637	113,979
$\begin{array}{c ccccc} 214,108 & 213,628 & 216,592 & 220,465 & 224,020 \\ \hline & 230,679 & 233,821 & 235,715 & 236,962 & 240,967 \\ \hline & 204,238 & 204,441 & 204,610 & 204,761 & 204,764 \\ \hline & 205,315 & 205,611 & 205,950 & 206,579 & 207,308 \\ \hline & 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ \hline & 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ \hline & 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ \hline & 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ \hline & 246,557 & 247,171 & 247,832 & 248,778 & 249,692 \\ \hline & 25,823 & 26,507 & 27,241 & 28,037 & 28,893 \\ \hline & 25,931 & 26,621 & 27,363 & 28,164 & 29,032 \\ \hline & 25,823 & 26,507 & 27,241 & 28,037 & 28,893 \\ \hline & 25,931 & 26,621 & 27,363 & 28,164 & 29,032 \\ \hline & 301,638 & 302,846 & 304,237 & 305,568 & 307,094 \\ \hline & any & 302,349 & 303,605 & 305,058 & 306,505 & 308,225 \\ \hline & 3,831,657 & 3,894,363 & 3,954,755 & 4,013,203 & 4,067,776 \\ \hline & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & y^{u} p p e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w e r \\ \hline & Forecasted & y^{1} o w$		214 109	215 (28	218 202	220.465	<u> </u>
Forecasted $y^{l}lower$ Forecasted $y^{l}lower$		214,100	213,626	210,392	220,465	224,020
Forecasted $\frac{204,200}{205,315}$ $\frac{204,411}{205,950}$ $\frac{204,701}{205,950}$ $\frac{204,701}{206,579}$ $\frac{207,308}{207,308}$ $\frac{246,557}{247,171}$ $\frac{247,832}{248,568}$ $\frac{249,692}{249,534}$ $\frac{250,632}{25,823}$ $\frac{26,507}{27,241}$ $\frac{28,037}{28,893}$ $\frac{25,931}{26,621}$ $\frac{27,363}{28,164}$ $\frac{29,032}{29,032}$ $\frac{301,638}{302,846}$ $\frac{304,237}{305,568}$ $\frac{307,094}{307,094}$ any $\frac{302,349}{303,605}$ $\frac{305,058}{305,058}$ $\frac{306,505}{308,225}$ $\frac{3,831,657}{3,894,363}$ $\frac{3,954,755}{4,013,203}$ $\frac{4,067,776}{4,067,776}$ $\frac{y^{2}lower}{y^{2}upper}$ Forecasted $\frac{y^{2}lower}{y^{2}upper}$ Forecasted $\frac{y^{2}lower}{y^{2}upper}$ Forecasted $\frac{y^{2}lower}{y^{2}upper}$ Forecasted $\frac{y^{2}lower}{y^{2}upper}$		204 238	204 441	204 610	204 761	204 764
Forecasted $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$ $y^2 lower$ $y^2 lower$ $y^2 lower$ Forecasted $y^2 lower$ $y^2 lower$		205.315	204,441	205,950	204,701	207.308
Forecasted 247,153 247,788 248,568 249,534 250,632 25,823 26,507 27,241 28,037 28,893 25,931 26,621 27,363 28,164 29,032 301,638 302,846 304,237 305,568 307,094 302,349 303,605 305,058 306,505 308,225 3,831,657 3,894,363 3,954,755 4,013,203 4,067,776 $y^{1}o w er$ $y^{u}p p er$ Forecasted $y^{1}o w er$ $y^{u}p p er$		246,557	247,171	247,832	248,778	249,692
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	Forecaste	d <u>247,153</u>	247,788	248,568	249,534	250,632
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$		25,823	26,507	27,241	28,037	28,893
$ \frac{301,638}{302,846} \frac{304,237}{305,568} \frac{307,094}{302,349} \frac{302,846}{303,605} \frac{304,237}{305,058} \frac{305,568}{306,505} \frac{308,225}{308,225} \frac{3,831,657}{3,894,363} \frac{3,954,755}{3,954,755} \frac{4,013,203}{4,067,776} \frac{4,067,776}{y^{1} o w e r} \frac{y^{1} o w e r}{y^{1} u p p e r} \frac{y^{1} o w e r}{y^{1} u p e r} \frac{y^{1} o w e r}{y^{1} u p e r} \frac{y^{1} o w e r}{y^{1} u p e r} \frac{y^{1} o w e r}{y^{1} u p e r} \frac{y^{1} o w e r}{y^{1} u p e r} \frac{y^{1} u v e r}{y^{1} u p e r} \frac{y^{1} u v e r}{y^{1} u p e r} \frac{y^{1} u v e r}{y^{1} u p e r} \frac{y^{1} u v e r}{y^{1} u p e r} \frac{y^{1} u v e r}{y^{1} u v e r} \frac{y^{1} u v e v e v e v e v e v e v e v e v e v$	ıda	25,931	26,621	27,363	28,164	29,032
any <u>$302,349$ $303,605$ $305,058$ $306,505$ $308,225$ 3,831,657 $3,894,363$ $3,954,755$ $4,013,203$ <math>4,067,776 $y^{1}ower$ $y^{u}pper$ Forecasted $y^{1}ower$ $y^{u}pper$ Forecasted $y^{1}ower$ $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$</math></u>	ce	301,638	302,846	304,237	305,568	307,094
$3,831,657 3,894,363 3,954,755 4,013,203 4,067,776$ $y^{1}ower$ $y^{u}pper$ Forecasted $y^{1}ower$ $y^{u}pper$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ $y^{1}ower$ Forecasted $y^{1}ower$ Forecasted $y^{1}ower$ Forecasted $y^{1}ower$ Forecasted $y^{1}ower$ Forecasted $y^{1}ower$ Forecasted $y^{1}ower$	nany		303,605	305,058	306,505	308,225
Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$ Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$ Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$ Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$ Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$ Forecasted $\frac{y^{l} o w e r}{y^{u} p p e r}$	n	3,831,657 y^lowen y^uppen	3,894,363 r	3,954,755	4,013,203	4,067,776
Forecasted y^{lower} y^{upper} Forecasted y^{lower} Forecasted y^{lower} y^{upper} Forecasted y^{lower} Forecasted y^{lower}		w^1 o ut o i				
Forecasted $y^l o w er$ $y^u p p er$ Forecasted $y^l o w er$ Forecasted $y^l o w er$ Forecasted $y^l o w er$	Forecaste	dyî bwer yî upper	r			
Forecasted $y^l o w e r$ $y^u p p e r$ Forecasted $y^l o w e r$	Forecaste Forecaste	d y lower y ûpper d y îlower y îupper	r - r			
Forecasted y ^l o w e r	Forecaste Forecaste Forecaste	dylower yûpper dyîlower yûpper dyîlower yûpper	r 			
y ^u p p e r	Forecaste Forecaste Forecaste Forecaste	dylower yûpper dylower yûpper dylower yûpper dylower yûpper	r			

3,837,258 3,900,328 3,961,460 4,021,477 4,077,922

The graphs of actual and predicted predictions are presented in Figure 1. It can be seen that the predicted values fit well with the actual values. The RMSE values also support this.



Figure 1. Cumulative graph and forecasted trend of countries **4. Conclusions**

The rapid spread of the Covid-19 pandemic among countries has prompted the need for research on infectious diseases' spread mechanism and predictions. This study aims to analyze Covid19 five-day prediction forecasts for G7 countries using the Prophet model, one of the machine learning models. For this purpose, five-day forecasts, five-day forecast intervals and RMSE statistics for the results obtained with the Prophet model were calculated.

Based on the forecasting results obtained using the Prophet model, Canada's closest predictions (lowest RMSE) were obtained. Canada and Germany, Italy, Japan, Italy, Japan and the UK have particularly close predictions for the first day. The possible reasons for the close predictions are that these countries have implemented nationwide shutdowns and did not change their data enough to affect the results during the 180 days. The most distant predictions (highest RMSE) are obtained for the USA. The reason for the highest RMSE for the USA is the parameter added to the model for shutdown days. In the USA, closure decisions were made at the state and county level rather than the federal government at the beginning of the pandemic and were inconsistent. For France, which has the highest RMSE results in predictions with the USA, the reason can be attributed to data corrections at the beginning of the pandemic. In light of these results, it is observed

that the prophet model as a machine learning model provides accurate predictions when the correct data is provided, the number of data-related corrections is reduced, and a regular shutdown regime is followed.

Forecasting the spread of COVID-19 is important for policymakers because it helps them to make informed decisions about how to respond to the pandemic. Accurate forecasts of the number of cases can inform decisions about lockdowns, school closures, and other public health measures. They can also help policymakers to plan for the distribution of vaccines and other medical resources. Additionally, forecasts can help policymakers to identify areas of the population that may be at particularly high risk, so that they can target interventions to those groups. Overall, the policy makers would consider the prophet model in forecasting the spread of infectious to limit the adverse effects on economies and businesses.

Funding: This research received no external funding.

Acknowledgments: The author is a PhD student at Institute of Graduate Programs, Ankara Haci Bayram Veli University. This article is partly derived from the PhD thesis and extended with additional analyses. Conflicts of Interest: The author declares no conflict of interest.

References

- Anirudh, A. (2020). Mathematical modeling and the transmission dynamics in predicting the Covid-19 What next in combating the pandemic. *Infectious Disease Modelling*. 5:366-74.
- Baldwin, R. (2020). The supply side matters: Guns versus butter, COVID-style. Retrieved from https://cepr.org/voxeu/columns/supply-side-matters-guns-versus-butter-covid-style
- Benvenuto, D., Marta G., Lazzaro V., Silvia A. and Massimo C., (2020). Application of the ARIMA Model on the COVID-2019 Epidemic Dataset. *Data in Brief*. 29:105340. doi: 10.1016/j.dib.2020.105340.
- Ceylan, Z. (2020). Estimation of COVID-19 Prevalence in Italy, Spain, and France. Science of The Total Environment. 729:138817. doi: 10.1016/j.scitotenv.2020.138817.
- Duman, A. (2020). COVID-19 ile Artan Eşitsizlikler ve Yoksulluk. Retrieved from https://sarkac.org/2020/06/covid19-ile-artanesitsizlikler-ve-yoksulluk/
- Furtado, P., (2021). Epidemiology SIR with Regression, Arima, and Prophet in Forecasting Covid-19. The 7th International conference on Time Series and Forecasting. *MDPI*.
- He, Z. and Tao, H. (2018). Epidemiology and ARIMA Model of Positive-Rate of Influenza Viruses among Children in Wuhan, China: A Nine-Year Retrospective Study. *International Journal of Infectious Diseases*. 74:61-70. doi: 10.1016/j.ijid.2018.07.003.
- IMF. (2021). World Economic Outlook.
- IMF. (2022). World Economic Outlook.
- Ioannidis, J. P. A., Sally C., and Martin A. T. (2022). Forecasting for COVID-19 Has Failed. *International Journal of Forecasting*. 38(2):423-38. doi: 10.1016/j.ijforecast.2020.08.004.
- Malik, M. I. (2020). Analysis and Forecast of COVID-19 Pandemic in Pakistan. medRxiv. doi: 10.1101/2020.06.24.20138800.
- Prajapati, S., Swaraj, A., Lalwani, R., Narwal, A. and Verma, K. (2021). Comparison of Traditional and Hybrid Time Series Models for Forecasting COVID-19 Cases. *arXiv.* doi: https://doi.org/10.48550/arXiv.2105.03266.
- Robson, W. (2019). The Math of Prophet. Retrieved from https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a.

- Roubini, N. (2020). A Greater Depression?. Retrieved from https://www.projectsyndicate.org/commentary/coronavirusgreater-great-depression-by-nouriel-roubini-2020-03
- Sah, S., Surendiran, B., Dhanalakshmi, R., Mohanty, S. N., Alenezi, F. and Polat, K. (2022). Forecasting COVID-19 Pandemic Using Prophet, ARIMA, and Hybrid Stacked LSTM-GRU Models in India. *Computational and Mathematical Methods in Medicine*. 2022:1
- Sardar, I., Muhammad A. A., Víctor L., Ahmed A. and Pradeep M. (2022). Machine Learning and Automatic ARIMA/Prophet Models-Based Forecasting of COVID-19: Methodology, Evaluation, and Case Study in SAARC Countries. *Stochastic Environmental Research and Risk Assessment*. doi: 10.1007/s00477-022-02307-x.
- Taylor, S. J. and Letham, B. (2018). Forecasting at Scale. *The American Statistician* 72(1):37-45. doi: 10.1080/00031305.2017.1380080.
- Viboud, C., Lone S., and Chowell, G. (2016). A Generalized-Growth Model to Characterize the Early Ascending Phase of Infectious Disease Outbreaks. *Epidemics*. 15:27-37. doi: 10.1016/j.epidem.2016.01.002.
- Voyvoda, E., and Yeldan, E. (2020). Salgın, Türkiye Ekonomisi ve Gerçekçi Bir Kamu Politikası Önerisi. Retrieved from https://sarkac.org/2020/06/salgin-turkiye-ekonomisi-ve-gercekci-bir-kamu-politikasionerisi/
- Wang, Y., Zehui Y., Wang, D., Yang, M., Li, Z., Gong, X., Wu, D., Zhai, L., Zhang, W. and Wang, Y. (2022). Prediction and Analysis of COVID-19 Daily New Cases and Cumulative Cases: Times Series Forecasting and Machine Learning Models. *BMC Infectious*
- WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19. Retrieved from https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-oncovid-19-11-march-2020

World Bank. (2020). Projected Poverty Impacts of COVID-19 (coronavirus).