Predicting COVID-19 incidences based on machine learning

Liudmyla Gryzun[1,*,†], Olexiy Lytovchenko [1,†]

[1] *Simon Kuznets Kharkiv National University of Economics, 9A Nauky Ave., Kharkiv, 61166, Ukraine*

Abstract

The paper is devoted to the important issues of predicting the trend of new cases of COVID-19 based on machine learning (ML), enabling preparation and adaptation to pandemic evolution. In the progress of work, based on the relevant theoretical framework, there was achieved the goal as for developing a web application for predicting COVID-19 incidence based on ML, and presenting the results of this work. There were undertaken the number of core steps. Some selected ML models for prediction were analyzed and estimated as for their accuracy in terms of various metrics. Based on these estimates there was proposed the Holt-Winters as the most appropriate one for the task of COVID-19 incidence forecasting. The said model was implemented as a mathematical basis for the development of authors' web application. The core stages of the application development are characterized. The functionality of the application is highlighted and analyzed. It is concluded that there were overcome some limitations of the similar applications revealed in the theoretical part of the paper. In the application it is realized global forecasting as well as for single country (region). Confidence interval for forecasting is suggested, which indicates the error of the model for a specific dataset. The modes of information visualization and representation are significantly widened. Data downloading is improved in terms of forming separate files of different formats. The prospects of the research are outlined in the lines of automatizing the choice of better ML model and adding the facility for the user to compare the models.

Keywords

machine learning, predicting models, web application, COVID-19 incidences prediction

1. Introduction

Since the emergence of the COVID-19 virus, humanity has faced many problems. These are primarily economic (slowdown of the production of essential goods, disruption of the product supply chain, losses in national and international businesses or their complete closure) and social spheres (implementation of social distance, as a result in the service sector, the impossibility of providing proper services and the closure of cultural, religious places and cancellation of activities related to entertainment).

IntelITSIS'2024: 5th International Workshop on Intelligent Information Technologies and Systems of Information Security, March 28, 2024, Khmelnytskyi, Ukraine

* Corresponding author.
† These authors contributed equally.
mailto:Lgr2007@ukr.net (L.Gryzun); mailto:lytovchenko.oleksij.o@hneu.net (O.Lytovchenko)
0000-0002-5274-5624 (L.Gryzun); 0009-0001-1739-8559 (O.Lytovchenko)

This problem is still relevant today, because the number of people infected with COVID-19 tends to rise and fall, despite the advent of a vaccine. After all, the virus is constantly mutating. There is public health consensus that vaccination is an effective prevention strategy. However, long-term investigations are necessary to estimate the clinical effects of vaccines and tests of their side effects on different groups of patients [1, 2]. In addition, different studies focusing on different time frames in pandemic trend prediction, came to the same conclusion as for high probability that the pandemic will remain a common disease or become endemic in the future. It means that the mankind has to learn to coexist with it.

The said evidence and socio-economic consequences of the COVID pandemic encourages the use of mathematical methods to analyze the epidemic evolution and to plan relevant response strategies accordingly [3]. At the same time, the rapid progress of artificial intelligence (AI) in the health care domain opens additional opportunities to social and medical experts. Machine learning (ML) as a branch of AI that embraces developing algorithms and mathematical models enabling computers to learn automatically from data without being explicitly programmed [3, 4], gives reliable and functional tools for forecasts building. ML is understood as a process of teaching computers to learn from data patterns and make decisions (predictions) based on the learning results. ML algorithms are designed to identify relationships in large datasets and implement them to build forecasts of new data behavior [3, 4, 5].

Under these conditions, the global community adapting to the changes and getting more digital, needs for relevant software tools which could assist health care structures in directing their efforts, and facilitate the social and business processes through predicting threatens and calculating risks caused by pandemic.

In this context, the development of an application that allows predicting the trend of new cases of COVID-19, enabling preparation and adaptation to them avoiding main risks and losing less profit is relevant. As mathematical tools when developing the application, it is relevant to use ML methods that allow building models for forecasting based on available data sets as for COVID incidence.

Several studies have already been carried out with the aim of predicting the evolution of COVID-19 pandemic in the world implementing ML methods. According to recent research, there are attempts to investigate the outcomes of forecasting based on different models [4, 5, 6].

However, most of the studies provide results for a specific region (country) instead of global ones [4]. In addition, it is pointed out that most models have a high risk of bias and significant concerns regarding their efficient applicability [3].

The authors' analysis of existing applications proving the pandemic predictions (given in details below) revealed their serious limitations in terms of both their technical facilities and accuracy of their forecasts.

Thus, despite the current theoretical and practical achievements in the lines of ML models investigations and applications building to obtain reliable COVID-19 incidence predictions for the countries all over the world, the topic of our research is relevant and important.

The goal of the work is to develop a web application for predicting COVID-19 incidence based on ML, and to present in the paper the progress and results of this work.

To achieve the said goal, there were undertaken the number of core steps that make main contribution of the paper. Firstly, there were analyzed recent studies as well as existing applications realizing similar functions of the subject domain so as to determine their shortcomings and limitations which reveal challenges of forecasting COVID-19 incidences based on ML models and make a theoretical background of our research.

Then, there was provided the investigation on purpose of the selection of the proper ML model for prediction: the set of ML models were tested to estimate their accuracy in terms of various metrics (RMSE, MAE, MAPE) and select the most appropriate one for the task of COVID-19 incidence forecasting.

Finally, the said investigation made a mathematical basis for the development of author's web-application on COVID-19 prediction building which was done. The stages of the application building were covered, its functionality was presented and analyzed in terms of revealed limitations of existing analogues.

All of these core steps are presented below.

2. Related works

As we mentioned above, there are several recent studies that apply AI techniques in diseases predictions. Among them, for instance, there are works (1) presenting an online ML health assessment system for metabolic syndrome and chronic diseases [7]; (2) utilizing multicenter data for development of scoring system for a liver disease mortality prediction [8]; (3) combining online COVID-19 data to train and evaluate five non–time series ML models in forecasting infection growth [9, 10]. These and other studies demonstrated that ML is relevant for evaluating disease trends and is able to provide medical authorities with information to be used to prevent pandemic outspread. There are also some research findings on COVID-19 AI forecasts and the usage of mobile sensor data to identify and control potential contacts [9, 11,12, 13]. However, most of these studies do not cover all over the world pandemic outspread, as they represent the results obtained only in a specific region or single country [6].

In this context, it is important that in the number of studies special focus is made on the a systematic review on the use of AI techniques for predicting the disease hospitalization and mortality based on primary and secondary data sources. The said research papers were published in the years 2019-2022. They mostly applied Random Forest model as one with the best performance and that were trained using groups of individuals sampled from inhabitants of European and non-European countries. According to the assessment with PROBAST, presented for instance in [3] the said models are characterized as those which demonstrate a high risk of bias and concerns regarding their successful applicability.

There is also the set of papers presenting the approaches built around compartmental models [19, 20, 21]. According to the researchers [18], the core problem of using compartmental models for an epidemic outspread modeling is the calculation of the average number of secondary infections caused by an infectious individual. In addition, the hypothesis as for homogeneousness of study population is not always true in real life. Therefore, to overcome the said difficulties, there were developed some models based on AI and linear regression [22, 23, 24 and others].

It is pointed out by the researchers that being a powerful tool for making forecasts, ML approach faces serious difficulties connected with detecting the model which is suitable to the existing data. Thus, there are works presenting different methods to identify the one that best models the evolution of the COVID-19 epidemic. In particular, there was provided a comparative study of exponential smoothing techniques, ARIMA model, and Poisson counting models to choose the best model for data from Chile. The results obtained suggest that the ARIMA model is the most suitable for predicting the number of confirmed cases of COVID-19 while for the number of deaths the smoothing techniques seem more suitable [25].

Some researchers appealed to nonlinear regression, exactly growth models, intending investigation of COVID-19 evolution. For instance, in [19] the authors focused on the four nonlinear models and came to the conclusion that the Gompertz model is the most suitable. However, it is necessary to admit the limitations of the study in terms of short term forecasts obtained with the help of the said model and mostly computational rather than theoretical character of the study.

In the context of our research it is also essential to analyze facilities of existing applications which realize similar functions of the subject domain on building reliable forecasts of COVID-19 incidences and visualizing their results for common users. The market analysis testified that so far, there are not many analogues software for forecasting in any subject domain, and in the field of forecasting pandemic evolutions. It is even felt a lack of such applications available for ordinary users.

In terms of our research there were three software products selected for the analysis: IHME, European COVID-19 Forecast Hub and Worldometer. Characterizing their common features, it is possible to conclude the following. Most of these analogues are web applications. To work with these tools, there are no restrictions on paid subscriptions, services or a registration in the web application. They are accessible to any user on the Internet to apply them and view the information they need. They allow to visualize some existing data, and obtain short period forecasts of COVID-19 outspread. The more detailed comparative characteristics of the said applications based on their functional facilities are summarized in the Table 1.

The held comparative analysis enables to reveal some benefits of the applications which are worth following and also limitations which are to be overcome at the design of our application.

In terms of the benefits, it is essential to provide statistics related to existing data, particularly, if a web application builds forecasting and enables to compare forecasting and existing data as well as monitoring the trend of cases.

In addition, all the analyzed analogues provide data about different regions, like world statistic and continents (Europe, Asia, Africa and so on), which helps analyze and observe information in general and predict a potential raising of cases in a certain country, if there is evidence of deterioration of the pandemic situation in the surrounding countries. It is also beneficial to have a facility to compare the frequency of cases from one country to another, which is not widely available in the applications.

Some analogues provide the facility of downloading data on COVID-19 incidences. This is relevant feature, if a web application suggests forecasting. Then a user can compare predictions and cases that may presumably appear, analyze the error of forecasting for regions (countries), and/or collect information about cases. However, the said analogues suggest downloading data about all countries and regions in one common file, which can be seen as a limitation. Presumably, a user would like to have an opportunity to download data regarding only a certain country (region).

Table 1
Comparative characteristics of the applications realizing similar functions of the subject domain

Functional facilities	IHME	European COVID-19 Forecast Hub	Worldo meter
Forecasting facility	+	+	-
Statistics on newly infected cases	+	+	+
Statistics on new deaths	+	+	+
Vaccination statistics	+	-	+
Displaying the total number of infected cases (deaths)	+	-	+
Displaying statistics on any country in the world	+	-	+
The facility to compare the frequency of cases from one country to another	+	-	-
Displaying statistics for certain world regions	+	-	+
Displaying statistics for the entire pandemic period	+	+	+
Graphs are scaled	-	+	-
Displaying a confidence interval for forecasting	+	-	-
Option of changing the type of graphics	-	-	+
Option of choosing the date of last forecast	+	+	-
Opportunity to choose a forecast model	-	+	-
Facility to view daily statistics	+	-	+
Facility to view weekly statistics	-	+	-
Application support	-	+	+
Dynamic geographic maps for viewing statistics	+	+	-
Facility to download some data	+	+	-

In terms of providing reliable forecasting, it is crucial to suggest a user choosing a confidence interval. It is evidently that prediction can not give 100% accurate results, so the confidence interval can demonstrate an approximate error and show how effective the prediction model is exactly for this or that data set. Thus, this facility has to be provided.

All the information has be available to be visualized in the form of tables, graphics and charts, which is only partially allowed by the analyzed applications, therefore also can be seen as a limitation to be overcome in the authors' application.

Thus, summarizing the analysis of the related papers and current applications realizing similar functions of the subject domain, we could conclude the following. Despite the diversity of research on the making predictions of pandemics evolution and their practical implementations, there are limitations which reveal challenges of forecasting COVID-19 incidences based on ML models. Among them, it is relevant to mention: (1) predictions based on limited data sets; (2) not global forecasting but only for single country (region); (3) not revealed confidence interval for forecasting; (4) limited modes of information visualization and data downloading; (5) absence of holistic approach to detect most suitable model for reliable and long term forecasts; (6) heavy dependence of the source databases and their updating, and others. These limitations highlight challenges of forecasting COVID-19 incidences based on ML models and confirm the urgency of the research in this field.

On the other hand, the held analysis makes a theoretical background for our research which is used below both for ML model building and for development of the web-application for COVID-19 incidences prediction.

3. Proposed model

Based on the approaches and the findings presented in the studies analyzed above, and minding the objectives of our work, it was provided the investigation on purpose of the selection of the most appropriate ML model which can become a mathematical basis for the web-application on COVID-19 prediction building.

As it was mentioned above, to make predictions in different subject domains, there can be used many models. For building forecasts of COVID-19 there were chosen some models to compare their effectiveness and accuracy by methods of finding an error. Also, these methods are different in the way of calculating the prediction. The selected models description is given below.

According to studies [14], linear regression is a simple and efficient algorithm used in ML for modelling and predicting numerical data. It is a type of regression analysis that models the relationship between a dependent variable (the target variable to be predicted) and one or more independent variables (predictor variables).

The Gompertz growth algorithm is a mathematical model that is used to describe the growth of populations, biological systems, and other phenomena that exhibit sigmoidal growth patterns.

Holt-Winters method, also known as triple exponential smoothing, is a popular time series forecasting method that is used to model and forecast data with trends and seasonal patterns [16, 17]. Overall, the Holt-Winters method has several advantages over other forecasting methods, including its ability to handle seasonality, adaptive smoothing, good fit to data, predictive power, ease of implementation, and availability of software.

Whether Holt-Winters method is better than other time series forecasting models depends on the specific data and context of the forecasting problem. The Holt-Winters method takes into account seasonality and trend. Moreover, this method can be also improved by estimating the parameters of the model, which can lead to better results than models with fixed parameters. It should be mentioned that in some studies, the Holt-Winters method has been shown to produce more accurate forecasts than other popular models, such as ARIMA, especially for data with seasonality [17].

In order to estimate the said models accuracy, there were considered different approaches: mean absolute percentage error (MAPE), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE) and other metrics. RMSE metric was chosen for our estimates as a leading metric, based on the following reasoning and according to [16, 17, 26]. Firstly, RMSE is particularly sensitive to large errors because it squares the errors before averaging them. This means that RMSE penalizes large errors more than smaller ones, which can be advantageous when forecasting COVID-19 cases or deaths, where underestimating or overestimating significantly can have serious implications. In addition, it is a versatile metric which can be suitable for all three models depicted above.

Secondly, in the context of COVID-19 forecasting, data ranges can vary widely between different regions (e.g., between countries or within different states of a country). RMSE's sensitivity to large errors can reveal discrepancies in model performance across these varied ranges, which might be less apparent with MAE or even MAPE, where the impact of the error magnitude may be misrepresented due to averaging or percentage calculations.

Thirdly, when optimizing models, the differentiable nature of RMSE makes it suitable for optimization algorithms, facilitating the fine-tuning of model parameters. This characteristic can lead to more efficient model improvement over iterations compared to MAE, which can have non-differentiable points, or MAPE, which can be undefined or infinite for data points close to zero [17].

Thus, RMSE metric was approved as a leading one to estimate the said models accuracy on purpose of selecting one which is the most appropriate for building the COVID-19 reliable and long-term forecast. Though, to understand the whole picture, other metrics (MAE and MAPE) were also used for assessing the accuracy of the models for building the COVID-19 forecast.

Below there are presented the process and the results of estimations.

As it was mentioned above, linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:

$$y = B_0 + B_1 X + e, \tag{1}$$

where y is the predicted value of the dependent variable for any given value of the independent variable (x); $B_0$is the intercept; B_1 is the regression coefficient; e is the error of the estimate.

For calculation there were taken two datasets with period in 2 months (or 56 days), where there is a small difference from day to day in the first dataset, and enormous difference in the second dataset. The prediction obtained with the help of the linear regression is given in the Figure 1. For estimating the error, the metrics RMSE, MAPE and MAE were used. After selecting the best coefficients according to the metrics, the error made: for RMSE - 162675, for MAPE - 0,29, and for MAE - 19485.

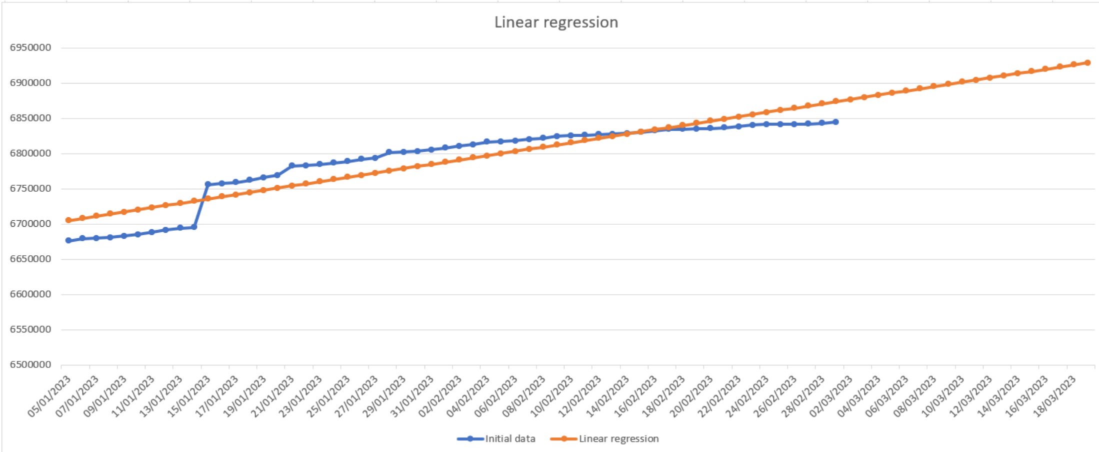


Figure 1: Prediction for a month by linear regression in the chart view for one dataset.

Let us consider other models and determine if the result can be improved.The nonlinear model is complex and, at the same time, is able to create more accurate results. The analysis develops a curve depicting the relationship between variables based on the dataset provided. As for nonlinear models, the most popular and accurate are considered to be Exponential, Gompertz, Verhulst, and Weibull growth models. These approaches are set up on constant growth of the data. According to the investigation [18], the most accurate model regarding COVID-19 forecasting is Gompertz growth. The formula of Gompertz function is as follows:

$$f(t) = ae^{-be^{-ct}}, \tag{2}$$

where a is an asymptote; b sets the displacement along the x-axis; c sets the growth rate (y scaling); e is Euler's Number. In other words, a, b and c are the coefficients that can be optimized to minimize the error and improve the accuracy of the model. In the image below it can be seen the chart with the prediction based on the Gompertz function (Figure 2): RMSE estimation is 73746, MAPE equals 0.085 and MAE is 5705. Now the accuracy of this model is better than for the linear forecast according to the metrics.

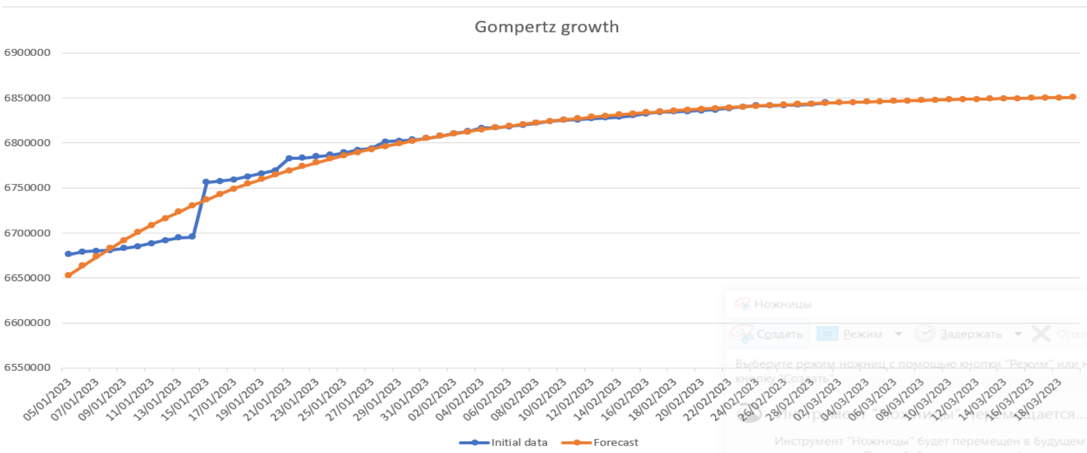


Figure 2: Prediction with help of the Gompertz function for the first dataset.

Then, the Gompertz model was tested on the second dataset with higher fluctuations in data, which revealed its drawbacks. For the second dataset RMSE is 2 630 431, MAPE equals 280,96 and MAE is 280 069, what makes the enormous error for the forecast. The chart can be seen in the Figure 3.

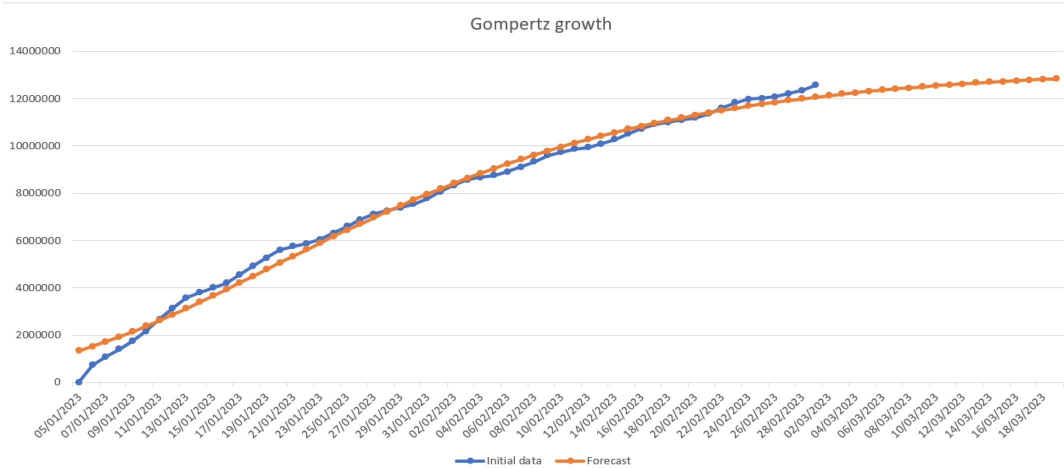


Figure 3: Prediction of cases by Gompertz growth on the second dataset with higher fluctuations in data.

Thus, it should be chosen a more effective and more sustainable model in terms of datasets with ono-smooth data.

The most popular and accurate traditional models are exponential smoothing and autoregressive integrated moving average (ARIMA) models [15, 17, 18]. According to the sources, they can be suitable for different situations, and in one case ARIMA model will be better, in another – exponential smoothing, but the difference in accuracy is not significant, and they can be alternatives to each other. In our case, for experimenting with data there was taken the exponential smoothing model.

Formula for calculating the forecast according to Holt-Winters method using exponential smoothing is as follows [15]:

$$F_{(t+1)} = (L_t + T_t)S_{(t-M+1)}, \tag{3}$$

where $L_t$is a level, T_t is a trend and $S_{(t-M+1)}$ is a seasonal factor.

Components of the formula can be calculated in such a way:

$$L_t = a\, Y_t/S_{t-M} + (1-a)(L_{t-1} + T_{t-1}), \tag{4}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}, \tag{5}$$

$$S_t = \Upsilon \frac{Y_t}{L_t} + (1-\Upsilon)S_{t-M}. \tag{6}$$

Here a, β and ϒ are the coefficients that can be optimized. In the pictures 7-8 it can be seen the results of calculations for exponential smoothing model and prediction (Figure 4).

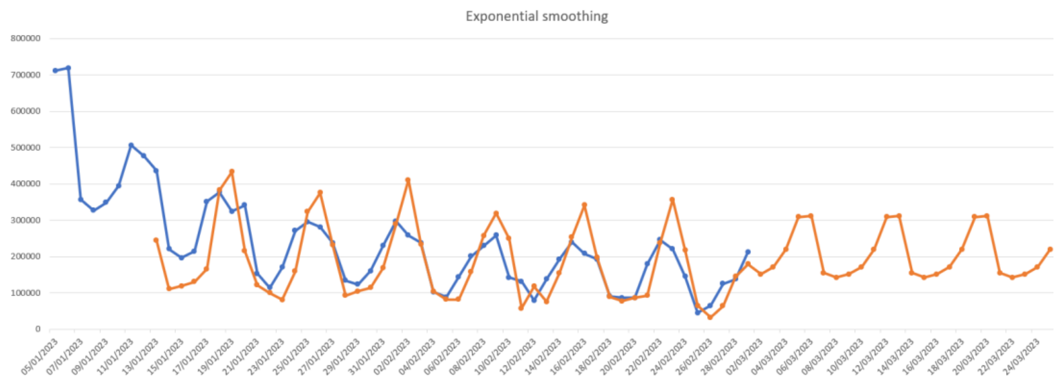


Figure 4: Prediction by exponential smoothing in the chart view.

As can be seen, now the RMSE for world cases (the second dataset) is 525 925, MAPE is 29,09 and MAE equals 57038. It means that in this situation exponential smoothing model is five times more effective than Gompertz function according to RMSE and MAE, and almost ten times more effective according to MAPE. Moreover, it is possible to obtain more accurate result adding more periods for teaching the model. The complete comparison of the models by different metrics and for different datasets is summarized in Tables 2-3.

Table 2
The comparison of linear regression and Gomperts function by different metrics for the first dataset.

Prediction model	RMSE	MAE	MAPE
Linear regression	162675	19485	0,29
Gompertz growth	73753	5705	0.085

Table 3
The comparison of Gomperts function and Holt-Winters method by different metrics for the second dataset.

Prediction model	RMSE	MAE	MAPE
Gompertz growth	2630431	280069	280,96
Holt-Winters	525925	57038	29,09

Thus, summarizing our estimates, we can conclude that the Holt-Winters method is a better choice than Gompertz growth and linear regression models for building predictions with a seasonal factor, based on the following reasons. Firstly, the Holt-Winters method is designed to handle both trend and seasonality, which are common in epidemiological data like COVID-19 case counts, whereas Gompertz growth and linear regression models do not take seasonality into account. Secondly, unlike simple linear regression which assumes a constant rate of change, or the Gompertz function which assumes a specific growth form, the Holt-Winters method can adapt to changes over time, providing more accurate short-term forecasts. In addition, the Holt-Winters method is more robust to outliers than simple linear regression. This is crucial for exactly COVID-19 data which can have sudden leaps due to changes in reporting practices. Finally, the Holt-Winters method has straightforward parameters (level, trend, and seasonality smoothing parameters) that are easy to interpret and adjust, unlike other models which may require more complex parameter tuning.

Therefore, the Holt-Winters model is selected as the most appropriate one for the task of COVID-19 incidence forecasting and implemented as a mathematical basis for the development of authors' web application.

4. Results

As a result, web-application for COVID-19 incidence forecasting based on ML learning was developed.

In the progress of the application development, database design was created. Additionally, restrictions were defined for this database in order to avoid violation and error in the software. The data dictionary was created to understand what tables and attributes the database contains and how it should be done pragmatically. The results of database design became conceptual, logical, physical models of the data.

As a result of the whole analysis, designing mathematical and database models the web application was developed using the stack of technologies. There were used Java language for the server side; HTML5/CSS3/JavaScript for the client side; PostgreSQL database for saving data; Spring, jQuery as frameworks and some additional libraries to ease development.

Characterizing the functional facilities of the developed web application, it is relevant to point out a number of features.

The developed web application provides COVID-19 data visualization, including predictions of new cases and deaths for different nations, continents and the world overall. Moreover, the application suggests confidence interval, which indicates the error of the model for a specific dataset (Figure 5).

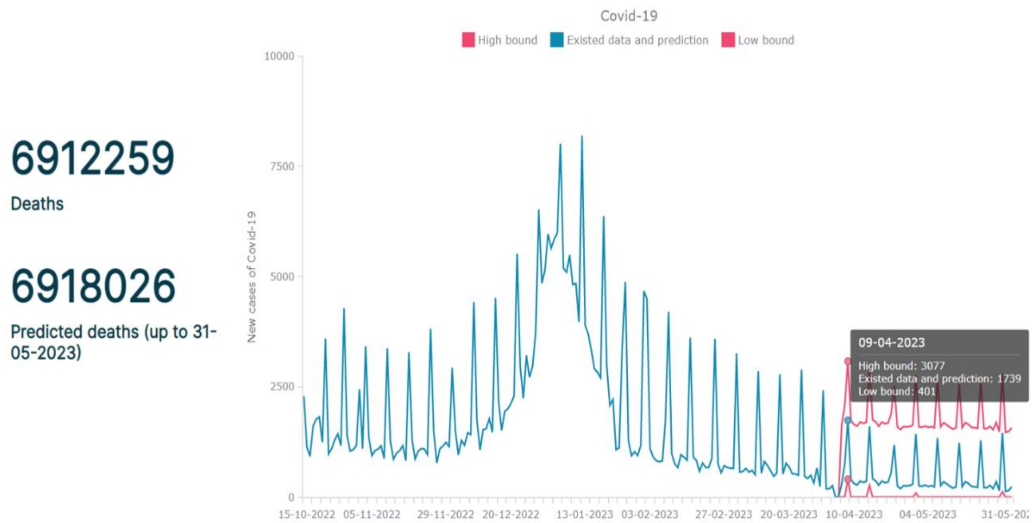


Figure 5: World statistics and demonstration of the confidence interval.

It visualizes data in different representations, and charts provide this data in various modes: cases of infection – deaths, new cases – total cases, daily – weekly etc. both in linear and area charts (Figures 6-7).

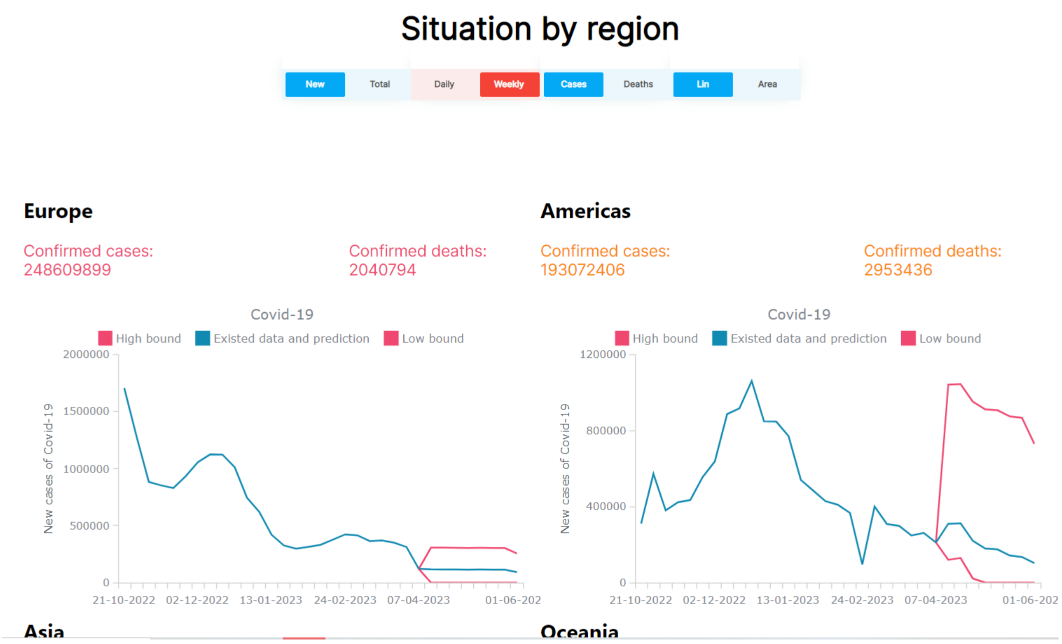
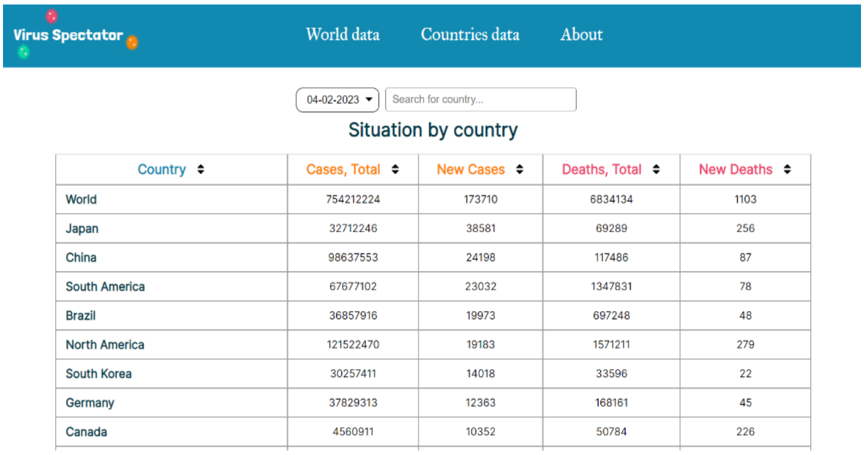


Figure 6: Region statistics in different modes.

Figure 7: Nation statistics in area chart mode.

The application also provides viewing statistics in the table form. Furthermore, it has additional functions as searching, sorting by different columns, and comparing data across different dates, offering a notable advantage to users (Figure 8).



Country	Cases, Total	New Cases	Deaths, Total	New Deaths
World	754212224	173710	6834134	1103
Japan	32712246	38581	69289	256
China	98637553	24198	117486	87
South America	67677102	23032	1347831	78
Brazil	36857916	19973	697248	48
North America	121522470	19183	1571211	279
South Korea	30257411	14018	33596	22
Germany	37829313	12363	168161	45
Canada	4560911	10352	50784	226

Figure 8: The table view of historical and predicted data.

Finally, the software provides opportunity to download data in different formats and dependently on the user need, it is possible to download data of any nation or region separately, without any spare information related to other countries in .csv or .json files for future comparison.

Coming from the depicted facilities, it is possible to conclude that there were overcome some limitations of the similar applications highlighted in the theoretical part of the paper. In particular, it is realized global forecasting as well as for single country (region). Confidence interval for forecasting is suggested, which indicates the error of the model for a specific dataset. The modes of information visualization and representation are significantly widened. Data downloading is improved in terms of forming separate files of different formats.

In the context of the prospects of the research, it would be beneficial to automatize the choice of better ML model and add the opportunity to compare the models for the user.

5. Conclusion

The paper is devoted to the important issues of predicting the trend of new cases of COVID-19 based on ML, enabling preparation and adaptation to pandemic evolution.

In the progress of work, based on the relevant theoretical framework, there was achieved the goal as for developing a web application for predicting COVID-19 incidence based on ML, and presenting the results of this work.

There were undertaken the number of core steps. Some selected ML models for prediction were analyzed and estimated as for their accuracy in terms of various metrics. Based on these estimates there was proposed the Holt-Winters as the most appropriate one for the task of COVID-19 incidence forecasting.

The said model was implemented as a mathematical basis for the development of authors' web application. The core stages of the application development are characterized. The functionality of the application is highlighted and analyzed.

It is concluded that there were overcome some limitations of the similar applications revealed in the theoretical part of the paper. In the application it is realized global forecasting as well as for single country (region). Confidence interval for forecasting is suggested, which indicates the error of the model for a specific dataset. The modes of information visualization and representation are significantly widened. Data downloading is improved in terms of forming separate files of different formats.

The prospects of the research are outlined in the lines of automatizing the choice of better ML model and adding the facility for the user to compare the models.

References

[1] N. Donthu, A. Gustafsson, Effects of COVID-19 on business and research, Journal of Business Research 117 (2020) 284-289. doi:10.1016/j.jbusres.2020.06.008.

[2] Y. Shang, H. Li, R. Zhang, Effects of Pandemic Outbreak on Economies: Evidence From Business History Context, 2021. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2021.632043/full

[3] S. Shakibfar, F. Nyberg, H. Li, J. Zhao, H. Nordeng, G. Sandve, M. Pavlovic, M. Hajiebrahimi, M. Andersen, M. Sessa M, Artificial intelligence-driven prediction of COVID-19-related hospitalization and death: a systematic review, Front. Public Health 11:1183725 (2023). doi: 10.3389/fpubh.2023.1183725.

[4] S. M. Salhout, Machine learning in healthcare strategic management: a systematic literature review, Arab Gulf Journal of Scientific Research (2023). doi: 10.1108/AGJSR-06-2023-0252.

[5] F. Saleem, A. Al-Ghamdi, M. Alassafi, S. AlGhamdi, Machine Learning, Deep Learning, and Mathematical Models to Analyze Forecasting and Epidemiology of COVID-19: A Systematic Literature Review, Int J Environ Res Public Health 19(9):5099 (2022). doi: 10.3390/ijerph19095099.

[6] C.S. Yu, S.S. Chang, T.H. Chang, J.L. Wu, Y.J. Lin, H.F. Chien, R.J. Chen, A COVID-19 Pandemic Artificial Intelligence-Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study, J Med Internet Res. 23(5):e27806 (2021). doi: 10.2196/27806.

[7] C.S. Yu, Y.J. Lin, C.H. Lin, Development of an online health care assessment for preventive medicine: A machine learning approach, J Med Internet Res, (2020). doi: 10.2196/18585.

[8] Y.J. Lin, C.H. Lin, S.T. Wang, S.Y. Lin, S.S. Chang, Noninvasive and convenient screening of metabolic syndrome using the controlled attenuation parameter technology: An evaluation based on self-paid health examination participants, J Clin Med (2019).

[9] S.M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, N. Kalhori, Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study, JMIR Public Health Surveill (2020). doi: 10.2196/18828.

[10] A. Yeung, F. Roewer-Despres, L. Rosella, F. Rudzicz Machine learning-based prediction of growth in confirmed COVID-19 infection cases in 114 countries using metrics of nonpharmaceutical interventions and cultural dimensions: Model development and validation, J Med Internet Res (2021). doi: 10.2196/26628.

[11] Z. Yang, Z. Zeng, K. Wang, S. Wong, W. Liang, M. Zanin et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J Thorac Dis (2020). doi: 10.21037/jtd.2020.02.64.

[12] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T. Hoang-Thi, S. Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia, Med Image Anal (2021). doi: 10.1016/j.media.2020.101860.

[13] L. Xu, R. Magar, F. Barati, A. Farimani, Forecasting COVID-19 new cases using deep learning methods, Comput Biol Med (2022). doi: 10.1016/j.compbiomed.2022.105342.
[14] M. Mondal, S. Bharati, P. Podder Diagnosis of COVID-19 Using Machine Learning and Deep Learning: A Review, Curr Med Imaging (2021). doi: 10.2174/1573405617666210713113439. PMID: 34259149.
[15] Time Series in Excel! Learn Exponential Smoothing Models for Time Series Forecasting in Excel, 2020. URL: https://www.analyticsvidhya.com/blog/2020/09/time-series-forecasting-ms-excel-exponential-smoothing/
[16] C. Bergmeir, R. Hyndman, B. Koo, A note on the validity of cross-validation for evaluating autoregressive time series prediction, Computational Statistics and Data Analysis 120 (2018) 70–83. doi:10.1016/j.csda.2017.11.003.
[17] Y.J. Mgale, Y.X. Yan, S. Timothy, A Comparative Study of ARIMA and Holt-Winters Exponential Smoothing Models for Rice Price Forecasting in Tanzania. Open Access Library Journal 8 (2021) 1-9. doi: 10.4236/oalib.110738.
[18] M. Atchadé, P. Tchanati, On computational analysis of nonlinear regression models addressing heteroscedasticity and autocorrelation issues: An application to COVID-19 data, Heliyon, 8 (10) (2022). doi:10.1016/j.heliyon.2022.e11057.
[19] Z. Yang, et al. Modified SEIR and AI prediction of the epidemics trend of Covid-19 in China under public health interventions, J. Thorac. Dis. 12 (2020) 165-174. doi: 10.21037/jtd.2020.02.64.
[20] O. Agossou, M.N. Atchadé, A. Djibril, Modeling the effects of preventive measures and vaccination on the Covid-19 spread in Benin republic with optimal control, Results Phys., 31 (2021). doi:10.1016/j.rinp.2021.104969.
[21] B.M. Ndiaye, L. Tendeng, D. Seck, Analysis of the Covid-19 pandemic by SIR model and machine learning technics for forecasting, Popul. Evol. (2020). doi:10.48550/arXiv.2004.01574.
[22] R. Sujath, J.M. Chatterjee, A.E. Hassanien, A machine learning forecasting model for Covid-19 pandemic in India Stoch. Environ. Res. Risk Assess. 34 (2020) 1-14. doi: 10.1007/s00477-020-01827-8.
[23] S.J. Fong, G. Li, N. Dey, R. Gonzalez-Crespo, E. Herrera-Viedma, Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak, Int. J. Interact. Multimed. Artif. Intell. 6 (2020) 132-140. doi:10.9781/ijimai.2020.02.002.
[24] S. Lalmuanawma, J. Hussain, L. Chhakchhuak, Applications of machine learning and artificial intelligence for Covid-19 (Sars-CoV-2) pandemic: a review Chaos Solitons Fractals, 139 (2020). doi:10.1016/j.chaos.2020.110059.
[25] C. Barría-Sandoval, G. Ferreira, K. Benz-Parra, P. López-Flores, Prediction of confirmed cases of and deaths caused by Covid-19 in Chile through time series techniques: a comparative study, PLoS ONE 16 (2021) 1-16. doi:10.1371/journal.pone.0245414.
[26] V. Prymak, B. Bartkiv, O. Holubnyk, Forecasting the exchange rate of the Ukrainian hryvna using machine learning methods Computer Systems and Information Technologies 1 (2023) 75–83. https://doi.org/10.31891/csit-2023-1-10.