Nonparametric method for estimation of forecasting models using small samples

Dmitriy Klyushin

Taras Shevchenko National University of Kyiv, prospect Glushkova 4D, Kyiv, 03680, Ukraine

Abstract
The coronavirus epidemic has stimulated a surge of research in the field of forecasting the epidemic curve based on various mathematical models. To predict the time series that predicts the number of patients, different models are used, both differential and machine learning models. Differential models for predicting the epidemic curve depend on a number of unpredictable factors. This often results in inaccurate predictions. In contrast to these approaches, machine learning models that predict time series based on training samples show higher reliability. In both cases, the problem arises of testing the hypothesis about the homogeneity of errors on training samples. The paper describes the application of the Klyushin-Petunin test to test the homogeneity of two samples and compares its effectiveness with the widely used the Wilcoxon test and the Diebold-Mariano test and the using the example of three methods for predicting the COVID-19 epidemic curve based on data on the number of cases in a certain period in Germany, Japan, South Korea and Ukraine. The efficiency and usefulness of the proposed nonparametric approach is demonstrated.

Keywords [1]
COVID-19, Machine Learning, Time Series, Forecasting, Nonparametric Test, Wilcoxon signed-rank test, Diebold–Mariano Test, Samples Homogeneity, Hill Assumption.

1. Introduction

The rapid and difficult to predict spreading of the coronavirus requires accurate and reliable forecasting of the epidemic curve. Classical differential models (SIR, SEIR, etc.) depend on many fuzzy unpredictable factors. This limits the use of differential models. As a rule, they provide short-term and scenario forecasts, allowing one to consider alternative scenarios for the development of events in the near future. This complicates the planning of measures to combat coronavirus and requires the use of other approaches. In opposite, machine learning methods have demonstrated good efficiency. They do not rely on preliminary assumptions on parameters of the epidemic curve, using only training samples. Using different forecasting models of the epidemic curve we must compare the distributions of forecasting errors to select the best model.

This paper describes a non-parametric approach to estimating errors in time series forecasting models using epidemic curve data from Germany, Japan, South Korea, and Ukraine as examples. Section 1 substantiates the relevance of the problem under consideration. Section 2 describes the COVID-19 epidemic curve prediction models. Section 3 describes common accuracy measures of time series forecasting and tests for their evaluation. Section 4 describes the Klyushin–Petunin test for samples without ties. Section 5 describes the Klyushin–Petunin test for samples with ties. Section 6 demonstrates the application of nonparametric methods using data from Germany, Japan, South Korea, and Ukraine. Section 7 contains conclusions and formulation of open questions. Section 8 contains the references.

2. Prognostic models of the epidemic curves

The predictive models used in forecasting the epidemic curve must be accurate and reliable [1], as decisions based on them have far-reaching implications. .

In epidemiology, differential compartmental models SIR, SEIR, and their modifications, are widely used. Model names reflect the compartments included in it: susceptible (S), infected (I), recovered (R), and exposed (E). They strongly depend on the parameters of transition from one compartment to another. SIR, SEIR models, as well as their modifications, are widely used to predict the COVID-19 epidemic curve [2–9]. They have a rigorous mathematical justification, but they depend on parameters that are difficult to accurately estimate. As a result, they are theoretically accurate, but practically very approximate models. Against this backdrop, methods of machine learning based on training samples often produce more precise predictions [10]

Prediction of the epidemic curve of COVID-19 in different countries is traditionally based on the use of SIR and SEIR models. In [2], the SIR model was applied to data from India. Similar work has been done by [3] in India based on a logistic curve and SIR models. The authors found that the logistic curve model is more precise than the SIR. In the paper of Guirao [4], the SIR and SEIR models have been simplified to reduce their dependence on weakly defined parameters. Ifghuis et al. [5] refined the SIR model for Morocco epidemic data. They compared the accuracy of the SIR and SEIR models, preferring the first model. All the models used in these works rely on the accuracy of the preliminary assumptions. If these assumptions are unreliable, the forecast becomes inaccurate. Some researchers have attempted to refine compartmental models using statistical methods. In the paper of Nesteruk [6], the optimal parameters for the SIR model are determined, the exact solution of the problem is obtained, and a short-term forecast of the epidemic curve in China is given. However, the precision of this model is also due to the reliability of the original data. In [7] and [8], the researchers proposed a modification of the SEIR model by adding environmental indicators to it. In [8] the authors tested the accuracy of this model on the example of China, Italy, South Korea and Iran. As in all other work using compartmental models, the authors point out that often the models depend on parameters that are difficult to measure accurately. In [9], the authors applied these models to predict infections in China, taking into account population mobility, latent infection rates, and virus infectivity.

Due to the unreliability of the assumptions underlying the compartment models, machine learning methods have been proposed as an alternative. In the article [11], Swapnarekha et al. compared statistical models, as well as machine and deep learning methods, using the COVID-19 epidemic curve prediction as an example. The authors have shown that the alternative methods they propose provide an accurate prediction of the epidemic. Sujat et al [12] used linear regression, multilayer perceptron and vector autoregression to predict the epidemic in India. It turned out that the multilayer perceptron allows us to get more accurate predictions, like linear regression and vector autoregression.

Appadu et al. [13] applied the Euler's iterative method and cubic spline interpolation to forecast the total number of infected people and the number of COVID-19 incidents in Germany, India, South Africa and South Korea. They compare the accuracy measures of the two forecasts using relative error. In the paper [14], Tuli et al. developed a cloud computing platform for their own mathematical model and tried to accurately predict the epidemic in real time. Distante et al. [15] simulated the Covid-19 outbreak in different regions of Italy using a deep convolution autoencoder and used the SEIR model to predict peaks in the epidemic curve using data collected in China for training. It turned out that these data allow for accurate predictions.

One of the most popular machine learning methods used to forecast the COVID-19 epidemic curve is an artificial neural network. Using recurrent neural networks, Kolozsvári et al. [16] predicted the COVID-19 epidemic curve and demonstrated its precision. Ibrahim et al. [17] to predict the epidemic curve proposed the LSTM-Autoencoder variational model, which includes epidemic and demographic data. The resulting model proved to be accurate both in the short and long term. Hu et al. [18] used artificial intelligence techniques to model the spread of the coronavirus. According to the authors, the models they propose give more precise predictions than standard algorithms. Guo and He [19]

proposed an artificial neural network that accurately predicts infection and mortality rates from COVID-19.

Balli [20] explored the application of the linear regression, the random forest method, the multilayer perceptron, and the support vector machine. It was shown that the support vector machine model is the most accurate. Kafieh et al. [23] forecast the epidemic curves in China, Germany, Iran, Italy, Japan, South Korea, Spain, Switzerland, and the United States. They used the multilayer perceptron, the random forest method, the recurrent neural networks, and the long short-term network. The last of these models turned out to be the most accurate. Rustam et al. [22] used the linear regression, the LASSO method, the support vector machine, and exponential smoothing to forecast the epidemic curve. The accuracy of all models used is high and comparable.

As we can see, various models are used to forecast the COVID-19 epidemic curve. Given this diversity, it is necessary to develop effective tools to compare their precision. It should be emphasized that a simple comparison of average precision rates is unreliable because it does not take into account its statistical nature. Traditionally, to assess the statistical validity of estimates of the precision of predictive models the nonparametric Wilcoxon test [23, 24] and parametric Diebold–Mariano test [25, 26] are used. In this work, we propose to use an original nonparametric test [27] to evaluate predictive models and demonstrate its effectiveness using three forecasting methods (the random forest method, the k-nearest neighbor method, and gradient boosting) for four countries: Germany, Japan, South Korea, and Ukraine.

3. Error measures

One of the most difficult forecasting tasks is selecting the most precise method. Model accuracy is usually assessed using standard error, mean absolute deviation, and mean absolute percentage error. However, before comparing mean accuracy scores, it is necessary to test the hypothesis of a statistically significant difference between their distributions. If the accuracy scores have the same distribution, the comparison of their means becomes unreliable, since the difference between the scores of the mean could be explained by chance.

Generally, model errors are assumed to be uniform, stationary, and unbiased. For example, errors are often considered to have a Gaussian distribution. In any case, it is necessary to test whether the accuracy estimates of the models have the same distribution. Traditionally, the nonparametric Wilcoxon test and the parametric Dibold-Mariano test are used to test this hypothesis. The null hypothesis states that the mean value of the difference between two samples of errors equals to zero.

As shown by numerical experiments, the Wilcoxon and Diebold-Mariano tests recognize location bias with a high level of confidence, but cannot recognize different variance at the same location with the same accuracy. The solution of this problem is the test proposed in [27]. This test recognizes both the shift in the mean and the change in the variance with a high level of confidence.

Since the samples of data containing information on errors of forecasting often are small (for example, in practice there are samples of the size 8 and even 6), we use resampling to increase the precision of statistical estimations. In particular, we use the bootstrap technique [28]. As well-known, this technique generates series samples that can contain duplicates (ties). That is why, to test homogeneity of two-samples of errors we need use two versions of the Klyushin–Petunin tests: without ties and with ties.

4. Nonparametric test for homogeneity of samples without ties

Let samples $u = (u_1, u_2, ..., u_n)$ and $v = (v_1, v_2, ..., v_m)$ obey absolutely continuous distribution functions F_1 and F_2, respectively. The null hypothesis on homogeneity is $F_1 = F_2$ and the opposite hypothesis is $F_1 \neq F_2$. According to Hill's assumption $A_{(n)}$ [29], if random values $u_1, u_2, ..., u_n$ obey exchangeable and absolutely continuous distribution function F then

$$P\left(u \in \left(u_{(i)}, u_{(j)} \right) \right) = \frac{j - i}{n + 1}, \qquad (1)$$

where $j > i$, u is a sample value obeying F, and $u_{(i)}$ and $u_{(j)}$ are the i-th and j-th order statistics.

Suppose that the null hypothesis is true and construct the variational series $u_{(1)}, u_{(2)}, ..., u_{(n)}$. Let $A_{ij}^{(k)} = \left\{ v_k \in \left(u_{(i)}, u_{(j)} \right) \right\}$. By Hill's assumption, if $j > i$ then

$$P\left(v_k \in \left(u_{(i)}, u_{(j)} \right) \right) = p_{ij} = \frac{j-i}{n+1}.$$

Construct the Wilson confidence interval for the probability of $A_{ij}^{(k)}$:

$$p_{ij}^{(1)} = \frac{h_{ij}^{(n,k)} n + g^2/2 - g\sqrt{h_{ij}^{(m,k)}(1 - h_{ij}^{(m,k)})n + g^2/4}}{n + g^2},$$

$$p_{ij}^{(2)} = \frac{h_{ij}^{(n,k)} n + g^2/2 + g\sqrt{h_{ij}^{(m,k)}(1 - h_{ij}^{(m,k)})n + g^2/4}}{n + g^2},$$

$$(2)$$

where $h_{ij}^{(n,k)} = \frac{\# A_{ij}^{(k)}}{n}$. Then, form the confidence interval $I_{ij}^{(n)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)} \right)$ whose significance level depends on g. If g equals 3 than the significance level of $I_{ij}^{(n)}$ is not greater than 0.05 [27]. Let $B = \left\{ p_{ij} = \frac{j-i}{n+1} \in I_{ij}^{(n)} \right\}$. Put $N = (n-1)n/2$ and find $L = \# B$. Then, $h = \frac{2L}{n(n-1)}$ is a measure of homogeneity of samples x and y (called p-statistics). Repeating the procedure of computing the Wilson confidence interval for the probability of B we obtain the interval $I^{(n)} = (p_1, p_2)$. Therefore, we may formulate the decision rule with the significance level, which is not greater than 0.05: if $I^{(n)}$ covers 0.95 then $F_1 = F_2$ else $F_1 \neq F_2$.

5. Nonparametric test for homogeneity of samples with ties

Let us extend the test described above to samples with ties, Let the sample $u = (u_1, u_2, ..., u_n)$ obeys distribution F. Denote the number of tie u_k in u by $t(u_k)$. A sample u without ties we shall call hypothetical, and a sample $\tilde{u}$ with ties we shall call empirical. If the distribution F is differentiable and Lipschitz-continuous, i.e.

$$|F(x) - F(y)| \leq K|x - y|$$

the sample value u^* is independent on u, and the order statistics $\tilde{u}_{(k)}$, $k \leq i$ of the empirical sample $\tilde{u}$ is a tie with multiplicity $t(\tilde{u}_k)$, then

$$p\left(\tilde{u}^* \in \left[\tilde{u}_{(k)}, \tilde{u}_{(k+1)} \right) \right) \approx \gamma\left(\tilde{u}_{(k)} \right) + \frac{1}{n+1}. \tag{4}$$

Therefore, if $p\left(\tilde{u}^* \in \left[\tilde{u}_{(i)}, \tilde{u}_{(j)} \right) \right)$, $i < j$, and $1 \leq i, j \leq m$, then

$$p_{ij} = p\left(A_{ij} \right) = p\left(\tilde{u}^* \in \left[\tilde{u}_{(i)}, \tilde{u}_{(j)} \right) \right) \approx \gamma_i + \gamma_{i+1} + ... + \gamma_{j-1} + \frac{j-i}{n+1}, \tag{5}$$

where $\gamma_l = \gamma\left(\tilde{u}_{(l)} \right)$, $A_{ij} = \left\{ \tilde{u}^* \in \left[\tilde{u}_{(i)}, \tilde{u}_{(j)} \right] \right\}$. If $\tilde{u}_{(l)}, i \leq l \leq j-1$ is not a tie, then $\gamma_l = 0$ and (5) is reduced to (1).

Denote by H the null hypothesis on homogeneity of absolutely continuous distribution functions F_1 and F_2. Let $\tilde{u}$ is a sample with ties. If $F_1 = F_2$ then the probability of $A_{ij}^{(k)} = \left\{ \tilde{u}_k \in \left(\tilde{u}_{(i)}, \tilde{u}_{(j)} \right) \right\}$ equals (5). Compute the Wilson confidence interval $I_{ij}^{(n,m)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)} \right)$ for $p\left(A_{ij}^{(k)} \right)$. Then

$$h^{(n,m)} = \frac{2 \# \left\{ p_{ij} = \dfrac{j-i}{n+1} \in I_{ij}^{(n,m)} \right\}}{n(n-1)}$$
is a homogeneity measure between empirical samples (called empirical p-statistics).

6. Forecasting of the COVID-19 epidemic curves

To demonstrate the usefulness of the proposed method, we consider models for predicting the COVID-19 epidemic curve based on data on the number of cases in a certain period in Germany, Japan, South Korea and Ukraine [10]. As the authors indicated [10], the selection of these countries was motivated by different character of pandemic dynamics in these countries and different anti-pandemic decisions made by their governments. This choice should mitigate the bias due to these factors. Note, that this is not unique variant of the countries selection. For example, Papastefanopoulos et al. [30] made a comparison of six time series methods to forecast percentage of active cases per population in ten other countries using the other criterion: the greatest number of total confirmed cases. Similar case studies were, for example, made in [31]. However, the choice of countries and forecasting methods is rather not important. The main is that we analyze error distributions of forecasting models applied the same datasets.
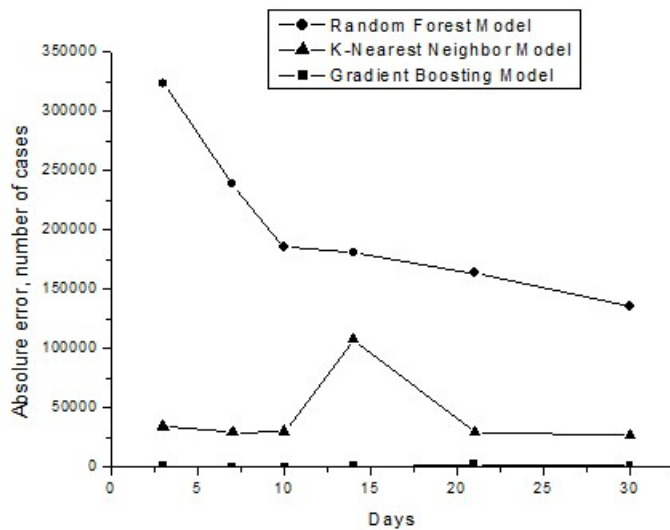

Figure 1: Absolute errors of forecasting models for Germany

Figure 1 demonstrates that absolute errors of the Gradient Boosting Model is varying about zero. The precision of the K-Nearest Neighbor Model is quite small also. The Random Forest Model is the worst model. Therefore, we could just compare these graphs and make decision. However, to complete analysis of the precision we must compare not the errors but the distributions of the errors. To do this we may use the two-sample tests, for example, the Wilcoxon sign-rank test, the Diebold–Mariano test and similar tests. But different tests have different features, for example, the Wilcoxon sign-rank test and the Diebold–Mariano text are good for comparison of samples with different means but they fail when samples have different variances and the same location. To solve this problem we can use the Klyushin–Petunin test [27]. The Wilcoxon sign-rank test, the Diebold–Mariano test and many other tests suppose that samples are homogeneous if they have the same average value (the hypothesis about the location shift).
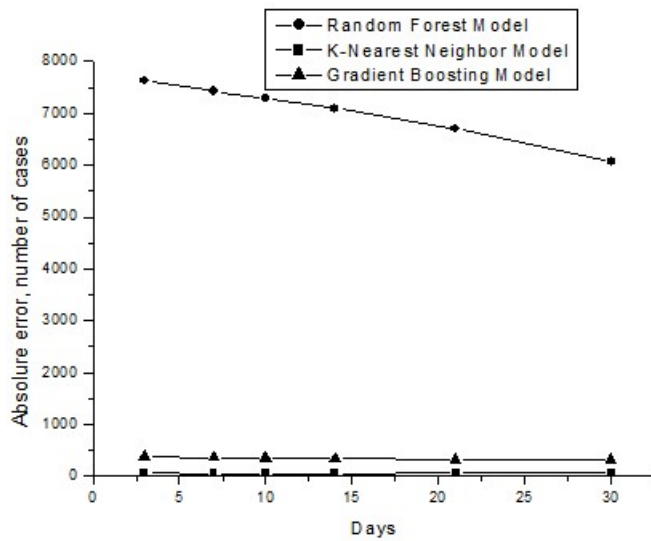
Figure 2: Absolute errors of forecasting models for Japan

As we see in Figure 2, the absolute errors of the K-Nearest Neighbor Model are smallest. The errors of the Gradient Boosting Model is varying about zero as in the previous case. The precision of the Random Forest Model is the largest.

It is quite interesting, that in the case of Japan we see that the Random Forest Model is the most imprecise, but the Gradient Boosting Model and the K-Nearest Neighbor Model changed their places. Nevertheless, analyzing the graphs demonstrated in Figure 2 we could suppose that the distributions of the errors of these models are equivalent.
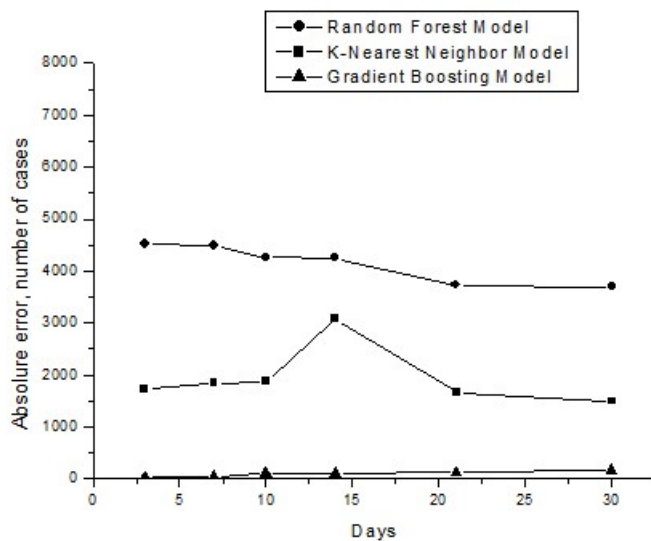

Figure 3: Absolute errors of forecasting models for South Korea

The graphs of the absolute errors of the Random Forest Model, K-Nearest Neighbor Model, and the Gradient Boosting Model for South Korea is similar to the results for Germany but the scales of the absolute errors are different. The K-Nearest Neighbor Model is the most accurate for the data from South Korea and vary different from the precision of KNN Model for Germany. As wee see in Figure 1 the absolute error of KNN Model for Germany is varying about 50000 cases, but for South Korea the range of the absolute error is bounded by the dozens of cases.
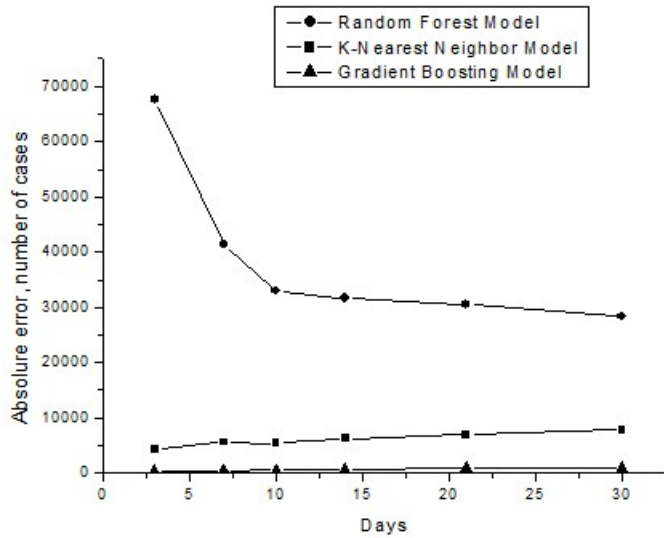
Figure 4: Absolute errors of forecasting models for Ukraine

The results of forecasting for Ukraine completely correspond to the results obtained for Germany, Japan, and South Korea. Is this case, we see that the absolute errors of the K-Nearest Neighbor Models and the Gradient Boosting Model are close. Therefore, to compare their precision correctly we must estimate their statistical equivalence. To compare statistical homogeneity of these data we apply the Klyushin–Petunin test, the Wilcoxon signed-rank test and the Diebold-Mariano test to real data from Germany, Japa, South Korea and Ukraine.

The main problem that can create an obstacle is the small size of the samples. To overcome this obstacle we must use some resampling technique applying the Klyushin-Petunin test. In this work we use the bootstrapping [28] using 10000 trials. We compare the absolute errors of the Random Forest Model (RFM), the K-Nearest Neighbor Model (KNN), and the Gradient Boosting Model (GBM).

Table 1

P-value of the tests for RFM, KNN and GBM methods for Germany

Tests/ Methods	Klyushin–Petunin			Diebold–Mariano			Wilcoxon		
	RFM	KNN	GBM	RFM	KNN	GBM	RFM	KNN	GBM
RFM	0.000	0.060	0.060	0.000	0.023	0.016	0.000	0.002	0.002
KNN	0.060	0.000	0.060	0.023	0.000	0.194	0.002	0.000	0.002
GBM	0.060	0.060	0.000	0.016	0.194	0.000	0.002	0.002	0.000

The p-values of the Klyushin-Petunin test and the Wilcoxon signed-rank test cited in Table 1 do not vary for all comparisons of the distributions of the absolute errors because the samples of the absolute errors corresponding to all the models are not pair-wise overlapping. This is explained by the fact that both these methods use ordering and permutations. That is why, for not overlapping samples they produce the constant p-value. Therefore, according to the Klyushin–Petunin test (the p-value is greater than 0.05) and the Wilcoxon signed-rank test (the p-value is less than 0.05) all the models for Germany provide statistically different results with the significance level $\alpha=0.05$. The Diebold-Mariano test recognizes the difference between all the samples with the exception of the K-Nearest Neighbor Model and the Gradient Boosting Model, as its p-value in all other cases is less that 0.05.

As we see, the forecasting errors of the Random Forest Model, K-Nearest Neighbor Model and the Gradient Boosting Model are very suitable for application of all the classical tests for samples homogeneity because these samples have clearly different locations. However, this is not the case when samples have similar locations but different scales.

Table 2
P-value of the tests for RFM, KNN and GBM methods for Japan

Tests/ Methods	Klyushin–Petunin			Diebold–Mariano			Wilcoxon		
	RFM	KNN	GBM	RFM	KNN	GBM	RFM	KNN	GBM
RFM	0.000	0.060	0.060	0.000	1.94e-05	1.95e-05	0.000	0.002	0.002
KNN	0.060	0.000	0.060	1.94e-05	0.000	1.24e-05	0.002	0.000	0.002
GBM	0.060	0.060	0.000	1.95e-05	1.24e-05	0.000	0.002	0.002	0.000

As in the previous case, the p-values of the Klyushin-Petunin test and the Wilcoxon signed-rank test cited in Table 2 do not vary for all comparisons of the distributions of the absolute errors because the samples of the absolute errors corresponding to all the models are not pair-wise overlapping. Therefore, according to the Klyushin–Petunin test and the Wilcoxon signed-rank test all the models for Japan provide statistically different results with the significance level α=0.05. The Diebold-Mariano test recognizes the statistical difference between all the samples.

Table 3
P-value of the tests for RFM, KNN and GBM methods for South Korea

Tests/ Methods	Klyushin–Petunin			Diebold–Mariano			Wilcoxon		
	RFM	KNN	GBM	RFM	KNN	GBM	RFM	KNN	GBM
RFM	0.000	0.060	0.060	0.000	2.37e-04	2.92e-05	0.000	0.002	0.002
KNN	0.060	0.000	0.060	2.37e-04	0.000	1.44e-02	0.002	0.000	0.002
GBM	0.060	0.060	0.000	2.92e-05	1.44e-02	0.000	0.002	0.002	0.000

Here, we see again, that the p-values of the Klyushin-Petunin test and the Wilcoxon signed-rank test cited in Table 3 do not vary for all comparisons of the distributions of the absolute errors because the samples of the absolute errors corresponding to all the models are not pair-wise overlapping. Therefore, according to the Klyushin–Petunin test and the Wilcoxon signed-rank test all the models for South Korea provide statistically different results with the significance level α=0.05. The Diebold-Mariano test recognizes the statistical difference between all the samples, also.

Table 4
P-value of the tests for RFM, KNN and GBM methods for Ukraine

Tests/ Methods	Klyushin–Petunin			Diebold–Mariano			Wilcoxon		
	RFM	KNN	GBM	RFM	KNN	GBM	RFM	KNN	GBM
RFM	0.000	0.060	0.060	0.000	0.039	0.036	0.000	0.002	0.002
KNN	0.060	0.000	0.060	0.039	0.000	0.002	0.002	0.000	0.002
GBM	0.060	0.060	0.000	0.036	0.002	0.000	0.002	0.002	0.000

Now, the p-values of the Klyushin-Petunin test and the Wilcoxon signed-rank test cited in Table 4 do not vary for all comparisons of the distributions of the absolute errors because the samples of the absolute errors corresponding to all the models are not pair-wise overlapping. Therefore, according to the Klyushin–Petunin test, the Diebold-Mariano test and the Wilcoxon signed-rank test all the models for Japan provide statistically different results with the significance level α=0.05.

As a general conclusion, we may state that all the tests recognize the statistical difference between the models. Therefore, we may rank the models by their precision. After statistical analysis, we may conclude that the Gradient Boosting Model is the best, the K-Nearest Neighbor Model is less accurate but quite good, and the Random Forest Models is the worst for forecasting the numbers of cases in Germany, Japan, South Korea, and Ukraine.

7. Conclusions and scope for the future work

Naïve comparison of the predicted values using error measures ignores the stochastic nature of these values. To solve this problem, we must make comparisons of the accuracy measures and arrange the models by their quality. Usually, for this purpose the nonparametric methods are widely used, in particular, the Wilcoxon signed-rank test and the Diebold-Mariano test. The experiments made using the data from Germany, Japan, South Korea, and Ukraine have shown that the Klyushin–Petunin test with bootstrapping as effective for comparing forecast models as the Wilcoxon signed-rank test and the Diebold–Mariano test.

The future work will be focused on the comparisons of the proposed test with other test used in this domain and on the theoretical properties of the p-statistics for the samples with unequal sizes.

8. References

[1] S. Eker, Validity and usefulness of COVID-19 models, Humanit Soc Sci Commun 54 (2020). doi:10.1057/s41599-020-00553-4.

[2] N. Anand, A. Sabarinath, S. Geetha, S. Somanath, Predicting the Spread of COVID-19 Using SIR Model Augmented to Incorporate Quarantine and Testing, Trans Indian Natl Acad Eng 5 (2020) 141–148, doi:10.1007/s41403-020-00151-5.

[3] M. Babu, M. Marimuthu, M. Joy, M. Nadaraj, E. Asirvatham, L. Jeyaseelan, Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models, Clin Epidemiol Glob Health 9 (2020) 26–33. doi:10.1016/j.cegh.2020.06.006.

[4] A. Guirao, The Covid-19 outbreak in Spain. A simple dynamics model, some lessons, and a theoretical framework for control response, Infect Dis Model. 5 (2020) 652–669, doi:10.1016/j.idm.2020.08.010.

[5] O. Ifguis, M Ghozlani, F. Ammou, A. Moutcine, A. Abdellah, Simulation of the Final Size of the Evolution Curve of Coronavirus Epidemic in Morocco using the SIR Model, J Environ Pub Health, 2020 (2020), article ID 9769267. doi:10.1155/2020/9769267

[6] I. Nesteruk, Statistics-based predictions of coronavirus epidemic spreading in mainland China, Innov Biosyst Bioeng 4 (2020) 13–18, doi:10.20535/ibb.2020.4.1.195074

[7] N. Wang, Y. Fu, H. Zhang, H. Shi, An evaluation of mathematical models for the outbreak of COVID-19, Precis Clin Med 3 (2020) 85–93. doi:10.1093/pcmedi/pbaa016.

[8] J. He, G. Chena, Y. Jiang, R. Jin, R. Shortridge, S. Agusti, M. He, J. Wu, D. Duarte, G. Christakos, Comparative infection modeling and control of COVID-19 transmission patterns in China, South Korea, Italy and Iran, Sci. Total Environ 74 (2020) 141447. doi:10.1016/j.scitotenv.2020.141447.

[9] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, B. Wei, L. Xin, X. Wei, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2), Science, 368 (2020) 489–493. doi:10.1111/jebm.12376.

[10] D. Chumachenko I. Meniailov, L. Bazilevych, T. Chumachenko, S. Yakovlev, Investigation of Statistical Machine Learning Models for COVID-19 Epidemic Process Simulation: Random Forest, K-Nearest Neighbors, Gradient Boosting, Computation 10 (2022) 86. https://doi.org/10.3390/computation10060086.

[11] H. Swapnarekha, H. Behera, J. Nayak, B. Naik, Role of Intelligent Computing in COVID-19 Prognosis: A State-of-the-Art Review, Chaos Solitons Fract, (2020) 109947. doi:10.1016/j.chaos.2020.109947.

[12] R. Sujath, J. Chatterjee, A. Hassanien, A machine learning forecasting model for COVID-19 pandemic in India, 34:959–972, Stoch Env Res Risk A. doi:10.1007/s00477-020-01827-8.

[13] A. Appadu, A. Kelil, Y. Tijani, Comparison of some forecasting methods for COVID-19, Alexandria Engineering Journal 60 (2020) 1565–1589. doi: 10.1016/j.aej.2020.11.011.

[14] S. Tuli, S. Tuli, R. Tuli, S. Gill, Predicting the Growth and Trend of COVID-19 Pandemic using Machine Learning and Cloud Computing. Internet Things, 11 (2020) 100222. doi:10.1016/j.iot.2020.100222.

[15] C. Distante I. Pereira, L. Gonçalves, P. Piscitelli, A. Miani, Forecasting Covid-19 Outbreak Progression in Italian Regions: A model based on neural network training from Chinese data, MedRxiv, 2020. https://www.medrxiv.org/content/10.1101/2020.04.09.20059055v1 doi:10.1101/2020.04.09.20059055.

[16] L. Kolozsvári, T. Bérczes, A. Hajdu, R. Gesztelyi, A. Tiba, I. Varga, G. Szőllősi, S. Harsányi, S. Garbóczy, J. Zsuga, Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence, MedRxiv, 2020. https://www.medrxiv.org/content/10.1101/2020.04.17.20069666v2. doi:10.1101/2020.04.17.20069666

[17] M. Ibrahim, J. Haworth, L. Lipani, A. Aslam, T. Cheng, N. Christie, (2020) Variational-LSTM Autoencoder to forecast the spread of coronavirus across the globe. MedRxiv, 2020. https://www.medrxiv.org/content/10.1101/2020.04.20.20070938v1. doi:10.1101/2020.04.20.20070938.

[18] Z. Hu, Q. Ge, S. Li, E. Boerwincle, L. Jin, M. Xiong, Forecasting and evaluating multiple interventions of Covid-19 worldwide. Front Artif Intell (2020) 2020.00041. doi:10.3389/frai.2020.00041

[19] Q. Guo, Z. He. Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence. Environ Sci Pollut Res 28 (2021) 11672–11682. doi:10.1007/s11356-020-11930-6

[20] S. Balli, Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. Chaos Solitons Fractals 142 (2021) 110512. doi: 10.1016/j.chaos.2020.110512.

[21] R. Kafieh, R. Arian, N. Saeedizadeh, Z. Amini, N. Serej, S. Minaee, S. Yadav, A. Vaezi, N. Rezaei, S. Javanmar, COVID-19 in Iran: Forecasting Pandemic Using Deep Learning. Computational and Mathematical Methods in Medicine (2021) article ID 6927985. doi:10.1155/2021/692798.

[22] F. Rustam et al. COVID-19 Future Forecasting Using Supervised Machine Learning Models. In: IEEE Access 8 (2020) 101489–101499. doi:10.1109/ACCESS.2020.2997311.

[23] B. Flores, The utilization of the Wilcoxon test to compare forecasting methods: A note, Int J Forecast 5 (1989) 529–535. doi:10.1016/0169-2070(89)90008-3.

[24] T. DelSole, M. Tippett, Comparing Forecast Skill. Mon Wea Rev, 142 (2014) 4658–4678. doi:10.1175/MWR-D-14-00045.1.

[25] F. Diebold, R. Mariano, Comparing predictive accuracy. J Bus Econ Stat 13 (1995) 253–263, doi: 10.1080/07350015.1995.10524599.

[26] F. Diebold, Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests, NBER Working Papers 18391, 2012, National Bureau of Economic Research, Inc.

[27] D. Klyushin, Y. Petunin, A Nonparametric Test for the Equivalence of Populations Based on a Measure of Proximity of Samples. Ukrainian Math J 55 (2003) 181–198. doi:10.1023/A:1025495727612.

[28] B. Efron, Bootstrap methods: another look on the jacknife. Ann. Statist. 7 (1979) 1–26. doi:10.1214/aos/1176344552

[29] B. Hill, Posterior distribution of percentiles: Bayes' theorem for sampling from a population. J Am Stat Assoc, 63 (1968) 677–691. doi:10.1080/01621459.1968.11009286

[30] V. Papastefanopoulos, P. Linardatos, S. Kotsiantis, COVID-19: a comparison of time series methods to forecast percentage of active cases per population. Applied sciences 10 (2020) 3880. doi:10.3390/app10113880.

[31] D. Klyushin, Comparing Predictive Accuracy of COVID-19 Prediction Models: A Case Study. In: Hassan, S.A., Mohamed, A.W., Alnowibet, K.A. (eds) Decision Sciences for COVID-19. International Series in Operations Research & Management Science, 2022, vol 320. Springer, Cham. doi:10.1007/978-3-030-87019-5_10.