10th International Young Scientists Conference on Computational Science

Baseline Accuracies of Forecasting COVID-19 Cases in Russian Regions on a Year in Retrospect Using Basic Statistical and Machine Learning Methods

A.V. Naumov[a], I.A. Moloshnikov[a], A.V. Serenko[a], A.G. Sboev[a,b,*], R.B. Rybka[a]

[a]NRC "Kurchatov Institute", Academician Kurchatov sq., 1, Moscow, 123098, Russia
[b]MEPhI National Research Nuclear University, Kashirskoye sh., 31, Moscow, 115409, Russia

Abstract

The large amount of data accumulated so far on the dynamics of the COVID-19 outbreak has allowed assessing the accuracy of forecasting methods in retrospect. This work compares several basic time series analysis methods, including machine learning methods, for forecasting the number of confirmed cases for some days ahead. Year-long data for all regions of Russia has been used from the Yandex DataLens platform. As a result, accuracy estimates for these basic methods have been obtained for Russian regions and Russia as a whole, in dependence on the forecasting horizon. The best basic models for forecasting for 14 days are exponential smoothing and ARIMA, with an error of 11–19% by the MAPE metric for the latest part of the course of the epidemic. The accuracies obtained can be considered as baselines for more complex prospective models.

Keywords: COVID forecasting; Machine learning; Time series analysis

1. Introduction

Forecasting the spread of the epidemic is an important task since it is necessary for planning mitigation policies, additional beds in hospitals, etc. Significant development in epidemic forecasting methods has been urged by the outbreak of the novel coronavirus infection (COVID-19). However, most of the works [7, 10, 12, 23, 24, 25] published in the first months of the pandemic strove to provide initial assessments of the evolving situation as early as possible. They were not aimed at validating the forecast accuracy in retrospect. Now that statistics for at least a year are available for most countries, it becomes more relevant to compare the performance of different forecasting methods [1, 2, 19, 20]. However, comparing the published scores of different approaches is complicated by the fact that different papers

use a wide variety of problem formulations (e.g., forecasting horizon), validation methods, and accuracy metrics. At the same time, only a few works (see review [4]) follow the good practice in time series forecasting [22] of validating the accuracy of the predictive method on known data. Moreover, different works rely on different amounts of available data.

The task of forecasting the numbers of confirmed cases [7, 19, 23], recoveries [7, 19] or deaths [1, 7, 19] is to predict, using input data on some previous days, the respective time series of interest for some days ahead. In the input data, each day is described by a vector that may, in addition to the past values of the time series, contain additional features indirectly characterising the course of events: mitigation measures in place [6, 21], a statistic of search queries containing keywords relevant to the epidemic [3], data inferred from social network texts [16]. The prediction depth is one day into the future [23, 25], seven days [9], or ten [1, 13, 19] or more days. Note the SberBank competition[1] of forecasting the number of confirmed cases in seven days in Russia and the Zindi competition[2] of forecasting the number of deaths from April 19 to June 8, 2020. Prediction is performed by a number of methods, including the following:

- the naive approach of considering the number of cases for a day equal to that number for the previous day (further referred to as dummy model) [2];
- exponential growth models [5];
- population models such as SIR [8, 10, 15, 17], SEIR [7, 10, 14, 17, 24, 25], and so on [1, 10, 17, 24];
- machine learning methods for regression [2, 9, 13, 19, 26], including SVM [19], LSTM [9, 20, 23, 25, 26], and deep neural network models [18].

The works that compare different forecasting approaches [2, 9, 19, 26] show that complex methods sometimes show no significant increase in accuracy compared to simplistic ones. For instance, complex machine learning methods are no better than simple ones at 10-day forecasting [19]. Due to that, it remains a relevant task to assess baseline accuracy levels achieved by basic forecasting methods on the data available to date, so that more complex methods that would be built in the future, possibly involving additional features within the input data, could be compared against said baselines.

Thus, the aims of this work are obtaining the baseline accuracy of forecasting the number of confirmed COVID-19 cases for Russia and studying the dependence of the accuracy on the forecasting range and on the amount of retrospective data used.

This will prepare the basis for further research and improvement of models for forecasting the dynamics of the epidemic.

2. Data

This work uses data from the Yandex DataLens[3] service for one year (from March 2020 to March 2021). The data contains daily numbers of confirmed cases, recoveries, and deaths for 85 regions of Russia. In addition to the regions, we consider Russia as a whole, obtained as the sum of the regions.

Data preprocessing included normalization and cleaning. In order to bring all regions to a single scale, the data were normalized per 100 000 population:

$$v_{\text{norm}} = v * \frac{100000}{P_{\text{region}}}, \tag{1}$$

where v is the value for the current day, P_{region} is the population size for the region. The population size for each region was taken from Rosstat data on the demography of the regions of the Russian Federation[4] (as of 01.01.2021) and considered constant throughout the period under consideration. Then the data was cleaned: all points before the

---

[1] https://ods.ai/competitions/sberbank-covid19-forecast/
[2] https://zindi.africa/competitions/predict-the-global-spread-of-covid-19
[3] https://datalens.yandex/covid19
[4] https://rosstat.gov.ru/folder/12781

day with the first detected cases were discarded, and, in order to clear the data from anomalies, all values less than or equal to 0 were replaced with interpolated values for neighboring, non-zero values.

All the three series – confirmed, recovered, and deceased – are highly dependent on each other. In this work, we will focus on predicting the number of cases because they are the closest to real-time indicators of the dynamics of the outbreak. We consider the forecasting problem as the prediction of the total number of new cases in i days ahead, i taking one of $1, 7, 14, 21, 28$. We have chosen to forecast the total increase in the number of confirmed cases over a certain period rather than separately for each day because we believe the former to be less dependent on random daily fluctuations caused by various external factors not directly related to the epidemic spread (delays in testing, in reporting, etc.), and therefore more stable. Technically, this means that the time series under forecasting is

$$v_{\text{sum}}(t) = \sum_{k=0}^{i-1} v_{\text{norm}}(t - k), \tag{2}$$

where the value for each day t is the sum of the normalized daily numbers of cases v_{norm} over i days. This new time series is i days less than the original one because it has i days excluded at the beginning.

3. Models

3.1. Dummy

The simplest basic model used in time series forecasting problems is "tomorrow as today". In this case, we take the current value of the time series as the predicted value. This approach can be used because we do not have any exceptional outliers and strong fluctuations in the series.

3.2. Statistical models

Analysis of a time series evolution is often done by alignment, smoothing, or applying various filters. These models account for trend and seasonality in the data by finding a functional dependence between the target prediction and the past values of the time series. We considered the exponential smoothing model (denoted as ES in Tables 1, 3, and 2) and the autoregressive model. For the exponential smoothing model, we used Holt Winter's method from the statsmodels library[5] with the "trend" parameter set as "additive", and other parameters kept default.

The autoregressive model assumes that the value of a process is linearly dependent on some previous values of the same process. We used AutoARIMA, the ARIMA model with automatic selection of optimal model parameters, from the pmdarima library[6].

The input data for statistical models is the entire time series, starting from the beginning of the training part of the current fold and up to the day of the testing part for which prediction is being made. Accordingly, for a later prediction date, a longer series is presented as the input to the model. The output is the value for a day i days ahead of the current day because, in the summed series, this corresponds to the total number of cases over i days from the current day.

3.3. Machine learning methods

The most popular machine learning methods often used to solve various regression problems include:

- Support Vector Machine with linear kernel and max_iter=5000 (LinearSVR);
- Linear Regression (LR): least-squares linear regression with normalize=True;
- Lasso: a linear least-squares model trained with L1 regularization, with max_iter=5000 and normalize=True;
- Ridge: linear least squares with L2 regularization, with normalize=True;
- Random Forest: an ensemble of decision tree models, with n_estimators=100;

---

[5] https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html
[6] https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.AutoARIMA.html#pmdarima.arima.AutoARIMA
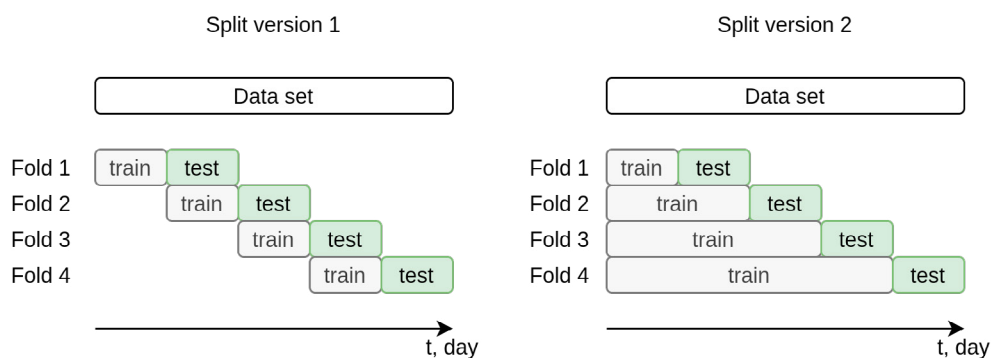
Split version 1   Split version 2

Fig. 1. Types of splitting the entire dataset into several folds of training parts, used for training models and fitting hyperparameters, and testing parts, used for evaluating the forecasting accuracy. Left: splitting version 1, non accumulating training part; right: splitting version 2, accumulating training part.

- Gradient Boosting with n_estimators=100.

All of the above models are taken from the sklearn library [11], with default parameters unless otherwise specified, and were used in the same way. As input data, the models are given a vector consisting of the values of the series for the last 14 days before the current day, while the output is the value of the series i days ahead.

4. Model scoring metric

The forecast accuracy is assessed by the mean absolute percentage error (MAPE) metric because, unlike MSE, it is normalized both by the length of the time series and by its scale, and, unlike R^2, it can be interpreted as a percentage: $\text{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$, where n is the number of days in the testing part, A_t and F_t are true and predicted values of the time series (normalized, preprocessed, and converted to summed values) for the day t.

5. Results

In order to evaluate each model at different stages of the epidemic (characterized, in particular, by different amounts of available past data), we split a time series for each region into five equal parts by time. Out of these parts, four pairs of training and testing parts, further referred to as folds, are formed as depicted in Figure 1. We consider two options for forming training and testing parts: accumulation of training data and a sliding window.

Table 1 compares of the forecasting models on different folds (train-test splitting performed with history accumulation) and with various forecasting horizons. For each fold and forecasting horizon, a model is characterized by its MAPE averaged over all regions. Separately, Russia as a whole is considered by summing the time series of all regions. Forecasting errors of all models on it are presented in Table 2. The true time series and the ones predicted by the dummy model and exponential smoothing are shown in Figure 2.

Below are the results that follow from Table 1:

- The error of various methods, even as simple as the dummy model, decreases as the epidemic progresses. This shows that the series become more predictable over time, even with simple models.
- For a short-term forecast (1–7 days), all models give almost the same accuracy as the basic approach of the dummy model. However, for a medium-term forecast, especially at later stages of the epidemic, exponential smoothing shows much lower error values than other models.
- Concerning a long-term forecast (28 days), the error for all models is too large to draw any conclusions based on the obtained forecast.

Table 1. MAPE (%), averaged over Russian regions, for different models, prediction horizons, and dataset parts (folds)

| Prediction range	Part №	Dummy model	AutoARIMA	ES	GBR	LR	Lasso	LinearSVR	RF	Ridge
1 day	1	14	15	19	48	36	59	30	44	39
	2	9	10	15	20	11	51	11	20	18
	3	4	4	7	32	6	64	5	30	13
	4	6	7	9	17	8	58	7	18	17
7 days	1	21	23	25	66	73	62	50	65	74
	2	14	14	13	40	20	45	16	39	30
	3	9	7	7	41	12	55	12	40	24
	4	12	10	10	29	12	29	12	29	24
14 days	1	30	33	35	67	123	90	97	67	105
	2	21	21	19	60	29	47	23	58	41
	3	18	12	11	46	23	49	20	46	37
	4	18	13	13	39	17	26	19	40	33
21 days	1	39	44	44	67	173	122	173	67	130
	2	27	28	27	71	39	52	31	72	50
	3	26	16	15	49	33	51	26	50	47
	4	23	17	15	40	22	33	26	41	39
28 days	1	46	56	53	61	184	130	262	61	140
	2	33	36	36	73	46	55	43	75	54
	3	34	21	20	51	43	53	31	52	53
	4	27	20	19	35	28	40	31	35	46
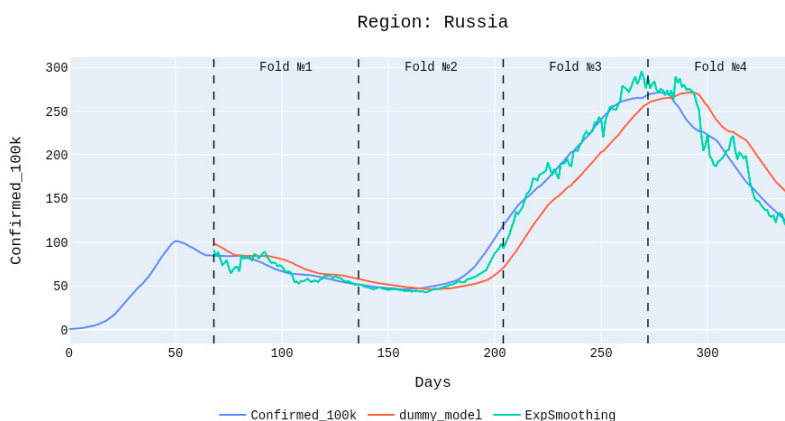
Fig. 2. True ("Confirmed_100k") and predicted ("Dummy_model" and "ExpSmoothing") series of the normalized 14-day total number of cases for Russia as a whole.

In Table 2, the best accuracy for different forecasting horizons in later periods of the epidemic has been achieved by ARIMA. The forecasting error for 14 days, averaged over all periods of the epidemic (folds), is 7%.

Overall, among machine learning models, linear regression shows the best accuracy for most regions. Of statistical methods, exponential smoothing shows accuracy either close or better than ARIMA but much more computational efficiency. We, therefore, pick one best performing model from each of the families – statistical and machine learning – and further show results for linear regression and exponential smoothing only.

Table 3 shows, for these two models, the effect of data accumulation (see the two splitting options in Figure 1). With non-accumulating (sliding window) training part (v1), certain folds are predicted better than others. However,

Table 2. The value of the MAPE error for different models for the Russia as a whole.

Prediction range	Part №	Dummy model	AutoARIMA	ES	GBR	LR	Lasso	LinearSVR	RF	Ridge
1 day	1	2	2	2	32	3	20	3	27	11
	2	2	2	2	11	2	28	2	9	10
	3	3	3	3	4	40	4	69	4	40	12
	4	3	3	3	18	3	40	3	19	13
7 days	1	5	4	4	38	39	11	25	39	25
	2	10	3	3	32	7	33	8	40	20
	3	8	3	3	57	7	68	6	57	28
	4	10	4	4	31	4	12	7	32	23
14 days	1	11	9	8	32	127	26	182	33	36
	2	15	7	8	63	16	38	9	64	30
	3	19	5	5	56	23	65	7	55	48
	4	17	8	7	38	6	16	13	38	34
21 days	1	17	12	13	29	117	31	315	30	34
	2	20	12	14	57	26	40	15	61	38
	3	29	7	7	61	38	64	23	61	59
	4	23	10	9	39	10	32	8	39	43
28 days	1	21	17	18	29	103	36	588	29	29
	2	22	17	18	57	27	39	18	57	38
	3	40	12	12	62	49	64	33	62	62
	4	26	13	14	32	22	45	18	32	45

Table 3. The mean MAPE (%) over Russian regions of two models, linear regression (LR) and exponential smoothing (ES), for 4 folds of the series, with non-accumulating training parts (v1) and with accumulating training parts (v2)

| Model | Prediction range | Part №1 | Part №2 | | Part №3 | | Part №4 | |
		v1,2	v1	v2	v1	v2	v1	v2
LR	1 day	36	14	11	11	6	14	8
	7 days	73	33	20	37	12	34	12
	14 days	123	50	29	90	23	55	17
	21 days	173	54	39	133	33	58	22
	28 days	184	61	46	249	43	73	28
ES	1 day	19	9	15	4	7	7	9
	7 days	25	13	13	7	7	10	10
	14 days	35	19	19	11	11	13	13
	21 days	44	26	27	15	15	15	15
	28 days	53	36	36	19	20	19	19

there is no clear tendency towards a decrease in the error value in predicting the later stages of the epidemic.

Accumulation of all past data into the training part (v2) has no noticeable effect on the accuracy of the exponential smoothing model. On the contrary, for machine learning models, the accumulation of training data plays a crucial role in the accurate forecasting of the epidemic.

The source code and results of experiments conducted in this work are available online[7].

---

[7] https://github.com/sag111/COVID-19-baselines-Russia

6. Discussion

This section analyses individual regions with the smallest and largest errors in predicting the total number of cases in 14 days. The plots show the exponential smoothing model because it performs better than others for this prediction range.

In Figures 3 and 4, the horizontal axis shows the number of days since the first case in the region, and the vertical axis shows the True ("Confirmed_100k") and predicted ("Dummy_model" and "ExpSmoothing") curves of the 14-day total number of cases, normalized per 100 thousand population.

The best accuracy has been achieved for Saratov, Volgograd, and Stavropol regions, with an average MAPE for all folds of 7%, 8%, and 8%, respectively. Prediction plots for the Saratov Region and Stavropol Krai are shown in Figure 3. It can be seen that in these regions, the curve of the epidemic was in itself not complicated: the first wave
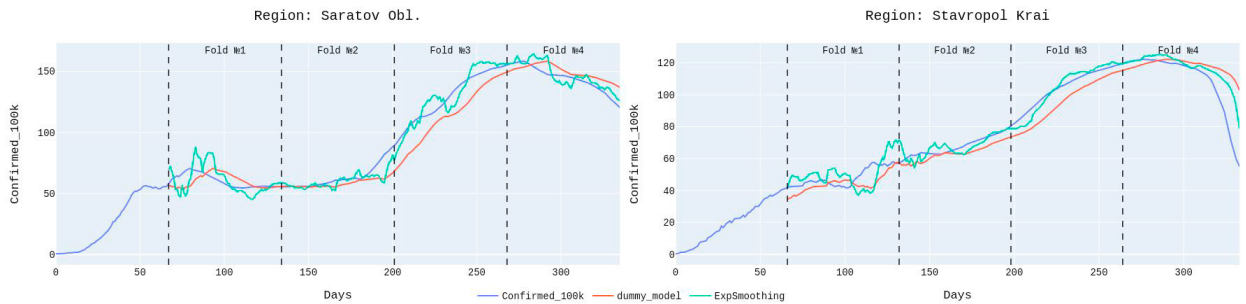

Fig. 3. True and predicted series of the 14-day total number of confirmed cases for regions with the highest prediction accuracy

of the epidemic was small and turned smoothly into the second one. This allowed exponential smoothing to perform well without significant errors at all stages of the epidemic.

The worst predicted regions were Tyva, Chukotka Autonomous Okrug (AO), and Nenets Autonomous Okrug, with MAPE of 46%, 91%, and 118%, respectively. Model prediction plots for the Chukotka AO and Nenets AO are shown in Figure 4. The significant forecasting error in these regions may be due to the small population of these regions, resulting in the high volatility of the data.
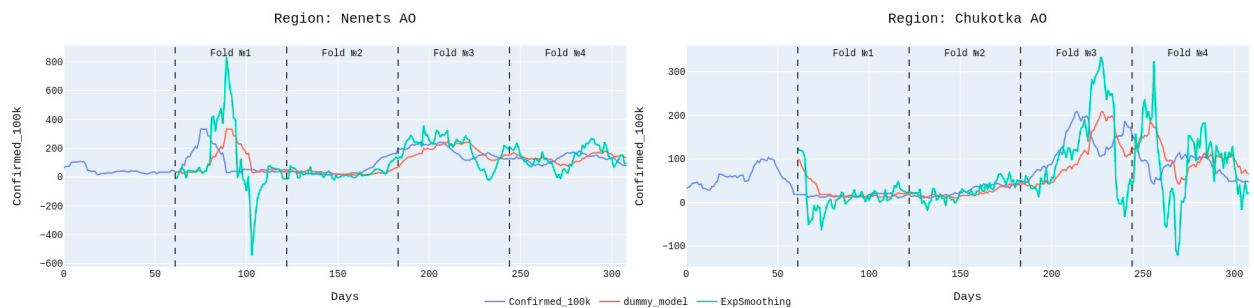

Fig. 4. True and predicted series of the 14-day total number of confirmed cases for regions with the lowest prediction accuracy

7. Conclusion

Within the family of machine learning models, logistic regression performs the best, showing high error at the beginning of the epidemic but improving as the data accumulates. Of the family of statistical models, the best results are achieved by exponential smoothing and ARIMA, but the former one is less computationally intensive. Overall,

forecasting the 14-day total number of confirmed cases in Russian regions is performed most accurately by logistic regression, exponential smoothing, and the naive (dummy) model, with the respective errors of 17%, 13%, and 18% by the MAPE metric.

The results obtained are a reference for the further progress of methods for predicting the dynamics of the coronavirus epidemic in Russia. Further work will consider:

- more advanced machine learning models, such as neural networks; population models such as SIR, SEIR, etc.;
- different approaches to enriching input data with additional features, such as weather data, self-isolation index, restrictive measures, features obtained from the analysis of social media such as Twitter (aspect-based sentiment analysis, the number of mentions of the disease, etc.);
- the possibility to build a single model for all regions of Russia.

Acknowledgements

References

[1] Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C., 2020. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLOS ONE 15, 1–21. URL: https://doi.org/10.1371/journal.pone.0230405, doi:10.1371/journal.pone.0230405.

[2] Chaurasia, V., Pal, S., 2020. Application of machine learning time series analysis for prediction covid-19 pandemic. Research on Biomedical Engineering , 1–13URL: https://link.springer.com/article/10.1007/s42600-020-00105-4.

[3] Derval, G., François-Lavet, V., Schaus, P., 2020. Nowcasting COVID-19 hospitalizations using Google Trends and LSTM. URL: https://crcs.seas.harvard.edu/files/crcs/files/ai4sg_2020_paper_57.pdf.

[4] Friedman, J., Liu, P., Troeger, C.E., Carter, A., Reiner, R.C., Barber, R.M., Collins, J., Lim, S.S., Pigott, D.M., Vos, T., Hay, S.I., Murray, C.J., Gakidou, E., 2020. Predictive performance of international covid-19 mortality forecasting models. medRxiv URL: https://www.medrxiv.org/content/early/2020/11/19/2020.07.13.20151233, doi:10.1101/2020.07.13.20151233.

[5] Grasselli, G., Pesenti, A., Cecconi, M., 2020. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. JAMA 323, 1545–1546. URL: https://jamanetwork.com/journals/jama/articlepdf/2763188/jama_grasselli_2020_vp_200049_1597263971.14383.pdf, doi:10.1001/jama.2020.4031.

[6] Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., Webster, S., 2020. Oxford COVID-19 government response tracker. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

[7] Hamzah, F.B., Lau, C., Nazri, H., Ligot, D., Lee, G., Tan, C., Shaib, M., Zaidon, U., Abdullah, A., Chung, M., et al., 2020. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Health Organ 1, 32. doi:http://dx.doi.org/10.2471/BLT.20.255695.

[8] Kröger, M., Schlickeiser, R., 2020. Analytical solution of the SIR-model for the temporal evolution of epidemics. part A: time-independent reproduction factor. Journal of Physics A: Mathematical and Theoretical 53, 505601. URL: https://doi.org/10.1088/1751-8121/abc65d, doi:10.1088/1751-8121/abc65d.

[9] İsmail Kırbaş, Sözen, A., Tuncer, A.D., Şinasi Kazancıoğlu, F., 2020. Comparative analysis and forecasting of covid-19 cases in various european countries with arima, narnn and lstm approaches. Chaos, Solitons & Fractals 138, 110015. URL: https://www.sciencedirect.com/science/article/pii/S0960077920304136, doi:https://doi.org/10.1016/j.chaos.2020.110015.

[10] Overton, C.E., Stage, H.B., Ahmad, S., Curran-Sebastian, J., Dark, P., Das, R., Fearon, E., Felton, T., Fyles, M., Gent, N., Hall, I., House, T., Lewkowicz, H., Pang, X., Pellis, L., Sawko, R., Ustianowski, A., Vekaria, B., Webb, L., 2020. Using statistics and mathematical modelling to understand infectious disease outbreaks: Covid-19 as an example. Infectious Disease Modelling 5, 409–441. URL: https://www.sciencedirect.com/science/article/pii/S2468042720300245, doi:https://doi.org/10.1016/j.idm.2020.06.008.

[11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[12] Perc, M., Gorišek Miksić, N., Slavinec, M., Stožer, A., 2020. Forecasting COVID-19. Frontiers in Physics 8, 127. URL: https://www.frontiersin.org/article/10.3389/fphy.2020.00127, doi:10.3389/fphy.2020.00127.

[13] Petropoulos, F., Makridakis, S., 2020. Forecasting the novel coronavirus COVID-19. PLOS ONE 15, 1–8. URL: https://doi.org/10.1371/journal.pone.0231236, doi:10.1371/journal.pone.0231236.

[14] Piovella, N., 2020. Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic. Chaos, Solitons & Fractals 140, 110243. URL: http://www.sciencedirect.com/science/article/pii/S0960077920306391, doi:https://doi.org/10.1016/j.chaos.2020.110243.

[15] Prodanov, D., 2020. Analytical parameter estimation of the SIR epidemic model. applications to the COVID-19 pandemic. arXiv preprint , arXiv:2010.07000.

[16] Qazi, U., Imran, M., Ofli, F., 2020. GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. SIGSPATIAL Special 12, 6–15.

[17] Roda, W.C., Varughese, M.B., Han, D., Li, M.Y., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? Infectious Disease Modelling 5, 271 – 281. URL: http://www.sciencedirect.com/science/article/pii/S2468042720300075, doi:https://doi.org/10.1016/j.idm.2020.03.001.

[18] Rodriguez, A., Tabassum, A., Cui, J., Xie, J., Ho, J., Agarwal, P., Adhikari, B., Prakash, B.A., 2020. Deep-covid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting. medRxiv URL: https://www.medrxiv.org/content/early/2020/09/29/2020.09.28.20203109, doi:10.1101/2020.09.28.20203109, arXiv:https://www.medrxiv.org/content/early/2020/09/29/2020.09.28.20203109.full.pdf.

[19] Rustam, F., Reshi, A.A., Mehmood, A., Ullah, S., On, B., Aslam, W., Choi, G.S., 2020. COVID-19 future forecasting using supervised machine learning models. IEEE Access 8, 101489–101499. doi:10.1109/ACCESS.2020.2997311.

[20] Shastri, S., Singh, K., Kumar, S., Kour, P., Mansotra, V., 2020. Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study. Chaos, Solitons & Fractals 140, 110227. URL: http://www.sciencedirect.com/science/article/pii/S0960077920306238, doi:https://doi.org/10.1016/j.chaos.2020.110227.

[21] Shvetsova, O., Adeel, A.B., Catalano, M., Catalano, O., Giannelli, F., Muftuoglu, E., Riggs, T., Sezgin, M.H., Tahir, N., VanDusky-Allen, J., Zhao, T., Zhirnov, A., 2020. Institutional origins of protective COVID-19 policies dataset. URL: https://orb.binghamton.edu/working_paper_series/7/.

[22] Tashman, L.J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16, 437–450. URL: https://www.sciencedirect.com/science/article/pii/S0169207000000650, doi:https://doi.org/10.1016/S0169-2070(00)00065-0. the M3-Competition.

[23] Wang, P., Zheng, X., Ai, G., Liu, D., Zhu, B., 2020. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. Chaos, Solitons & Fractals 140, 110214. URL: http://www.sciencedirect.com/science/article/pii/S096007792030610X, doi:https://doi.org/10.1016/j.chaos.2020.110214.

[24] Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 395, 689–697.

[25] Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., et al., 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease 12, 165. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7139011/, doi:10.21037/jtd.2020.02.64.

[26] Zeroual, A., Harrou, F., Dairi, A., Sun, Y., 2020. Deep learning methods for forecasting COVID-19 time-series data: A comparative study. Chaos, Solitons & Fractals 140, 110121. URL: http://www.sciencedirect.com/science/article/pii/S096007792030518X, doi:https://doi.org/10.1016/j.chaos.2020.110121.