Recommendation Rules Mining for Reducing the Spread of COVID-19 Cases

Vitaliy Yakovyna[a,b], Natalya Shakhovska[b], Khrystyna Shakhovska[b] and Jaime Campos[c]

[a] *University of Warmia and Mazury in Olsztyn, 2 Michała Oczapowskiego str., Olsztyn, 10-719, Poland*
[b] *Lviv Polytechnic National University, 12 Bandera str., Lviv, 79013, Ukraine*
[c] *Linnaeus University, PG Vejdes 6 & 7, Växjö, SE-35195, Sweden*

Abstract

The COVID-19 pandemic is having an unprecedented impact on society and the economy, affecting virtually every aspect of people's daily life and all sectors of the economy. In this situation, society and the health care system need help from modern technologies such as Artificial Intelligence, Big Data, and Machine Learning, which intended to help governments to choose and implement an adequate strategy to combat the spread of the disease by balancing between human safety and the constraints of social and economic life. This paper considers the recommendation rules extracted from the novel ensemble of machine learning methods such as regression tree and clustering. The merged Oxford COVID-19 Government Response Tracker and European Centre for Disease Prevention and Control Covid-19 Cases datasets have been used with the data ranged from January 01 to October 04, 2020. The conclusions and findings of the study could be helpful for decision making on appropriated state policy for reducing the spread of new Covid-19 cases.

Keywords[1]

COVID-19, Machine Learning, Regression Tree, Clustering, Prediction

1. Introduction

A new species of coronavirus (COVID-19) was first discovered in late 2019 in Wuhan, China and has since spread almost worldwide. Currently, the total number of reports of the spread of this viral disease covers 215 countries and territories [1]. Due to the rapid spread of cases, their wide territorial coverage and potentially serious consequences that threaten human life, the World Health Organization has declared the spread of COVID-19 a pandemic [2]. As of October 26, 2020, 43,694,475 confirmed cases of infection were identified, which led to 1,162,999 deaths [1].

The pandemic is having an unprecedented impact on society and the economy, affecting virtually every aspect of people's daily life and all sectors of the economy. According to some estimates, about 4 billion people, or about half the world's population, suffer from severe restrictions on mobility and restrictions on social communications. The COVID-19 pandemic has triggered a global economic crisis that is different from the global economic crises of previous decades and may have more severe and lasting consequences. Epidemiological constraints have led to significant changes in people's daily lives, their habits, social and economic ties, methods and forms of communication, production processes, and social and political life.

Due to such a significant impact on day-to-day people's lives and the world economy, the most important issues of concern to society and governments include: (i) when the spread of COVID-19 will reach its maximum, both at national and worldwide scales; (ii) how long the pandemic will last;

IDDM'2020: 3rd International Conference on Informatics & Data-Driven Medicine, November 19–21, 2020, Växjö, Sweden
EMAIL: yakovyna@matman.uwm.edu.pl (V. Yakovyna); nataliya.b.shakhovska@lpnu.ua (N. Shakhovska); kristin.shakhovska@gmail.com (K. Shakhovska); jaime.campos@lnu.se (J. Campos)
ORCID: 0000-0003-0133-8591 (V. Yakovyna); 0000-0002-6875-8534 (N. Shakhovska); 0000-0002-9914-229X (K. Shakhovska); 0000-0001-7048-8089 (J. Campos)

CEUR Workshop Proceedings (CEUR-WS.org)

(iii) how many people in general will be affected by the disease [3]. The most reliable and reasoned answers to such questions will help governments to choose and implement an adequate strategy to combat the spread of the disease by balancing between human safety and the constraints of social and economic life.

In this situation, society and the health care system need help from modern technologies such as Artificial Intelligence (AI), Big Data, and Machine Learning to fight new diseases and make rapid progress. Governments and health organizations need effective decision support systems during a pandemic to effectively control the virus and make optimal real-time decisions to curb the spread of the disease. Artificial intelligence systems and methods can also provide significant assistance in the development of the COVID-19 vaccine. AI methods and algorithms are used to analyze the spread of the disease, predict and track patients and potential carriers of infection. Intelligent analysis of data on confirmed cases of illness, recovery and death is carried out.

At present, the vaccine against COVID-19 is still undergoing clinical trials, and there is no proven effective cure for this disease. In such circumstances, the only effective way to slow the spread of infection is "social distancing" and the use of antiseptics and personal protective equipment to prevent the transmission of the virus from person to person. At the same time, the role of modeling and predicting the spread of the disease, the time and duration of the maximum number of infections and the scale of the epidemic for each country is growing. The findings of such modeling and forecasting are the basis for sound public health decisions regarding appropriate measures and the allocation of resources, such as pulmonary ventilation systems or additional field hospitals.

Behavior analysis of COVID-19 requires the development of a powerful mathematical apparatus for tracking, including automated, its dissemination to develop timely, dynamic, and sound solutions. As shown above, the methods and tools of artificial intelligence and machine learning are an adequate and effective basis for such tasks. The current COVID-19 pandemic poses serious challenges for data scientists and AI professionals due to limited and incomplete data and the large amount of heterogeneous data, as well as the significant impact of various factors on disease behavior and spread.

The purpose of the paper is to build the recommendation rules for appropriated state policy for reducing the spread of new Covid-19 cases. The system of these rules is based on novel ensemble of machine learning methods such as regression tree and clustering.

2. Related Works

The significant impact of the COVID-19 pandemic on all spheres of public life has led researchers to a significant interest in modeling the spread, behavior and origin of the virus, understanding its nature and properties, and finding ways to control the disease. Since the beginning of 2020, tens of thousands of articles have been published in various fields of science on the COVID-19 pandemic. Research institutions, foundations and governments are investing significant human, financial and technical resources to gain new knowledge about the pandemic and close gaps in understanding the nature of the disease and its consequences, including medical, social, and economic. Thus, a large-scale study and analysis of publications on COVID-19 using machine learning methods by A. Doanvo et al. [4] showed that the main volume of articles falls on research in the field of public health, pandemic, clinical diagnosis and treatment of coronavirus. At the same time, a small amount of work has been published on the microbiological details of this virus, including its pathogenesis and routes of transmission. A detailed review of the literature on the application of machine and in-depth learning algorithms in the processing of medical images for the diagnosis of coronavirus disease was conducted in [5] together with a comparative analysis of the application of each algorithm and its performance.

The development of new effective models for predicting the number and dynamics of new COVID-19 infections, as well as epidemiological forecasting in general, is extremely important for the health care system, as it enables effective planning to eliminate or reduce possible epidemics. The main requirement for such models is the maximum possible accuracy and reliability of the forecast of epidemiological time series. To meet these requirements, a number of AI-based models have been used for a number of years [6–8]. The methods used to solve the problem of constructing the COVID-

19 prediction model have ranged from statistical autoregression like ARIMA to more robust machine learning methods [3, 9–11, 13–19].

Thus, Ribiero et al. [9] consider different regression-based models like cubist, ridge, reference vector regressions, ARIMA along with random forest, and ensemble overlay training for one, three, and six days prediction of cumulative value of COVID-19 cases in Brazil. They demonstrate that, in most cases, support vector regression and stack ensemble training perform better against the accepted criteria than the other models studied. In general, the models developed in [9] can generate accurate forecasts with an error of less than 6.90%. In [9], it was concluded that further improvement in the development of COVID-19 prediction models can be achieved by combining deep learning and ensemble learning accumulation, by using multi-objective optimization for tuning the hyperparameters of prediction models and adopting a set of features that allows to explain the dependencies of future COVID-19 cases.

Zhang et al. [3] used a segmented Poisson model for prediction and analysis the time-series of the COVID-19 cases in six G7 countries. In the developed model, they included government interventions (advice/orders regarding staying at home, social distancing, blocking, and quarantine) as factors influencing the COVID-19 outbreak. The analysis makes it possible [3] to forecast statistically the tipping point, duration, and intensity of attacks for the countries under study. The paper [3] implies that if there were no serious coercive measures against control strategies like blocking, distancing, orders to stay at home, then the virus would spread exponentially. This indicates that the interventions/actions have significantly reduced the size of outbreaks and flattened the epidemic curves.

The dynamics of the COVID-2019 outbreak in China, Italy and France in January–March 2020 is analyzed in [10]. Some universal epidemic spread was revealed based on the analysis of simple daily lag maps, that allows the authors [10[to suggest that such simple models can be successfully used to quantify the spread of the epidemic, in particular the height and peak times of infected persons. The authors [10] carried out the analysis of the same dataset using a susceptible-infected-recovered-deaths model, and conclude that the parameter describing the patient recovery rate appears to be almost the same in all countries studied, while mortality and infection rates seems to be more volatile. However, Fanelli and Piazza [10] showed that it is necessary to reduce the infection rate sharply and quickly in order to see a noticeable decrease in the mortality rate and epidemic peak position.

Roosa et al. [11] exploited phenomenological models tested during previous outbreaks to evaluate short-term predictions of the total number of confirmed COVID-19 cases in Hubei Province, and separately for all other territory of China. They provide forecasts for 5, 10 and 15 days for five consecutive days, from 5 to 9 February 2020, with quantitative uncertainty based on the generalized model of logistical growth, the Richards growth model and the subepidemic wave model.

Vaishya et al. [12] conducted a detailed analysis of publications on the use of AI for the COVID-19 pandemic. The main AI applications in the COVID-19 pandemic have been shown to include, but not be limited to: diagnosis and early detection, patient treatment monitoring, tracking of infected persons, disease and mortality prediction, drug and vaccine development, reducing the burden on health workers and, finally, disease prevention. Thus, it is concluded in [12] that AI-based systems can be effectively used for early diagnosis and treatment of patients, in decision support systems in the medical field, as well as for training physicians and medical students on the COVID-19 disease, which will have an impact on the strategy and tactics of treatment of subsequent patients, and thus reduce the burden on medical staff in hospitals during a pandemic. AI technology can be useful for tracking potential ways of spreading the virus and informing the public about the threat of the disease through tracking the contacts of people with a confirmed diagnosis, including social platforms and networks. Among other things, machine learning models can predict the number and dynamics of COVID-19 incidence by region, age, social and other populations, thus helping to identify the most vulnerable areas and segments of the population. With the ability to analyze and process data in real time, AI can provide timely and up-to-date information to help prevent further spread of the infection. It can be used to predict the dynamics and direction of the disease, the dynamics of the number of diagnosed and severe cases in each region, to predict in advance the need for hospital beds, staff, medication during the crisis. In general, AI will play an increasing role in medicine and medical applications by providing proactive and prognostic measures in the field of health care [12].

Basu and Campbell [13] propose a model based on Long Short-Term Memory, which was trained on cumulative number of COVID-19 cases and deaths separately. They train the developed model using the dataset containing number of confirmed COVID-19 cases during the period of more than four months. The parameters of the proposed model can be adjusted in such a way as to provide the required forecasting accuracy. In the article [13] using the developed LSTM-model, a quantitative assessment and analysis of the impact of measures taken by various US counties on the spread of the disease has been carried out.

Neher et al. [14] consider a seasonal model based on a susceptible-infectious-recovered approach to analyze and predict the spread of COVID-19 virus reinfection in the world in the coming years. In their model, they consider such factors as the rate of infection, the volume and speed of emigration and population movement, and so on. An updated version of their model [15] aims to analyze and forecast the necessary hospital resources to combat pandemics of such a scale as COVID-19.

Hu et al. [16] built and investigated a machine learning model based on composite autoencoders to describe the spread of coronavirus in China in April 2020 based on data from previous periods. They clustered cities and provinces based on the characteristics obtained from the developed auto-coding model. Similar LSTM-based models have been used to study patient-related statistics [17], as well as to identify trends in the spread of the epidemic in China [18].

The study by S. Tuli et al. [19] is based on an improved machine learning model for analyzing and predicting the spread of the COVID-19 epidemic in different countries. The authors of [19] have developed and studied the Robust Weibull model with iterative weighting. They conclude that the developed Weibull model is statistically better than the Gaussian baseline model for prediction of the COVID-19 outbreak. It was shown [19] that simple Gaussian model results in an overly optimistic COVID-19 spreading scenario. The developed model was deployed on a cloud computing platform using the FogBus framework for more accurate and realistic forecasting of the epidemic's growth dynamics.

3. Material and Methods

3.1. Dataset Description

In this study we have used data from the Oxford COVID-19 Government Response Tracker as well as European Centre for Disease Prevention and Control (ECDC) Covid-19 Cases [20].

The Oxford COVID-19 Government Response Tracker (OxCGRT) dataset provides information on which governments took action, what action they did, and when they took it. The OxCGRT systematically gathers information on a range of common policy responses by a given government and determines the extent to which the government is implementing these measures. The respective scores are combined into a set of policy indicators.

The second dataset contains new public data on the geography of COVID-19 cases worldwide from European Centre for Disease Prevention and Control. Each line or record contains the number of new cases per day, by country or region.

For the performed analysis we have merged the mentioned datasets and considered data from January 01, to October 04.

3.2. The Methods Used

Stochastic gradient enhancement is a solution to a regression problem by building an ensemble of "weak" predictive decision trees. At the first iteration, a decision tree with a limited number of nodes is built. The difference between that predicted tree multiplied by the learning rate and the desired variable at that stage is then calculated. And the next iteration builds on this difference. This continues until the result stops improving. At every step we try to fix the errors of the previous tree.

3.2.1. Data Preprocessing

Dataset consists of attributes with different nature. An approach to processing large, distilled data was described in [21]. The noise data and outliers are presented in the dataset studied as well. That is why at the first stage the following steps are required:

- feature selection,
- empty data analysis,
- data normalization and scaling.

Feature selection is made based on theory of information. The joint mutual information between each feature and target attribute ConfirmedCases is calculated as:

$$JI(f_i) = \sum_{f_j \in S} \left(I(f_i;c) + I(f_i;c|f_j) \right)$$

where c is target class, $I(f,c)$ is mutual information, f_j is already selected feature, f_i is processed feature.

The already selected featured were selected manually. The list of the already selected features looks like the following: countryName, loaddate. As result, 19 from 38 features were selected.

The next step is empty data analysis. The dataset consists of 66,998 rows, 26,003 of them have empty values in selected attributes. Due to small quantity of rows with empty data these rows were eliminated. In addition, binning was made as well.

It is desirable to bring all input variables to a single range and normalize (the maximum absolute value of input variables should not exceed one). Otherwise, errors due to variables varying over a wide range will be more influential than errors due to variables varying over a narrow range. By ensuring that each feature changes within the same range, we ensure that each has an equal effect. Therefore, the input variables, as a rule, are scaled so that the variables change in the range of the function, as a rule, [0,1] or [−1,1]. The Softmax scaling is used. The distribution of dataset for each feature is shown in the Figure 1.

4. Results and Findings

First, the regression decision tree is developed. Decision trees divide an object space according to a set of splitting rules. These rules are logical statements about a variable and can be true or false. Three circumstances are key here:

1. the rules make it possible to implement sequential dichotomous data segmentation,
2. two objects are considered similar if they appear in the same segment of the partition,
3. at each step of the partition, the amount of information about the variable under investigation (response) increases.

The main feature of the propose algorithm is its k-arc structure. Branching on a chosen trait x splits training objects on k subsamples, where k is the number of different characteristic values.

Without loss of generality, we will assume that the feature x has values from {0, 1, …, k-1}, $k \geq 2$. In this case, when constructing of the decision tree, from the vertex x there are k arcs labeled with numbers from {0, 1, …, k-1}. Let σ be the label of one of the arcs leaving the vertex x, $\sigma \in \{0, 1, \ldots, k\text{-}1\}$. To form a new current subset of objects and a new current set of features, those objects from T are deleted for which the value of feature x is not equal to σ, and also feature x itself is removed from the set of features.

Let v be a hanging vertex generated by a branch of a tree with the inner vertices $x_{j_1}, \ldots, x_{j_r}$ and let the arc outgoing from the vertex x_{j_i}, $i \in \{1, \ldots, r\}$, be labeled σ_i. Further let $T(v)$ -the current set of objects that hit the vertex v. The vertex v is associated with a pair $(B, w(v))$, where $w(v)$ is equal to the mean value of the target variable over all objects from $T(v)$, and B is an elementary conjunction of the form $x_{j_1}^{\sigma_1} \ldots x_{j_r}^{\sigma_r}$. If the vertex v is not pendant, then we assign to it the conjunction $B = xx_{j_1}^{\sigma_1} \ldots x_{j_r}^{\sigma_r}$ ... Truth interval of elementary conjunction B denote by NB. Let S be a recognizable object. For each hanging vertex $(B, w(v))$, a check is performed that the description of the test object belongs to the truth interval NB. If the description S belongs to NB, then the object S is associated with the value of the target variable $w(v)$. Object S is assigned the value of the target variable

$$W = \frac{\sum_{i=1}^{l} w(v_i) * I_{B_i}}{\sum_{i=1}^{l} I_{B_i}},$$

where

$$I_{B_i} = \begin{cases} 1, if\ S \in N_{B_i}, \\ 0, otherwise. \end{cases}$$

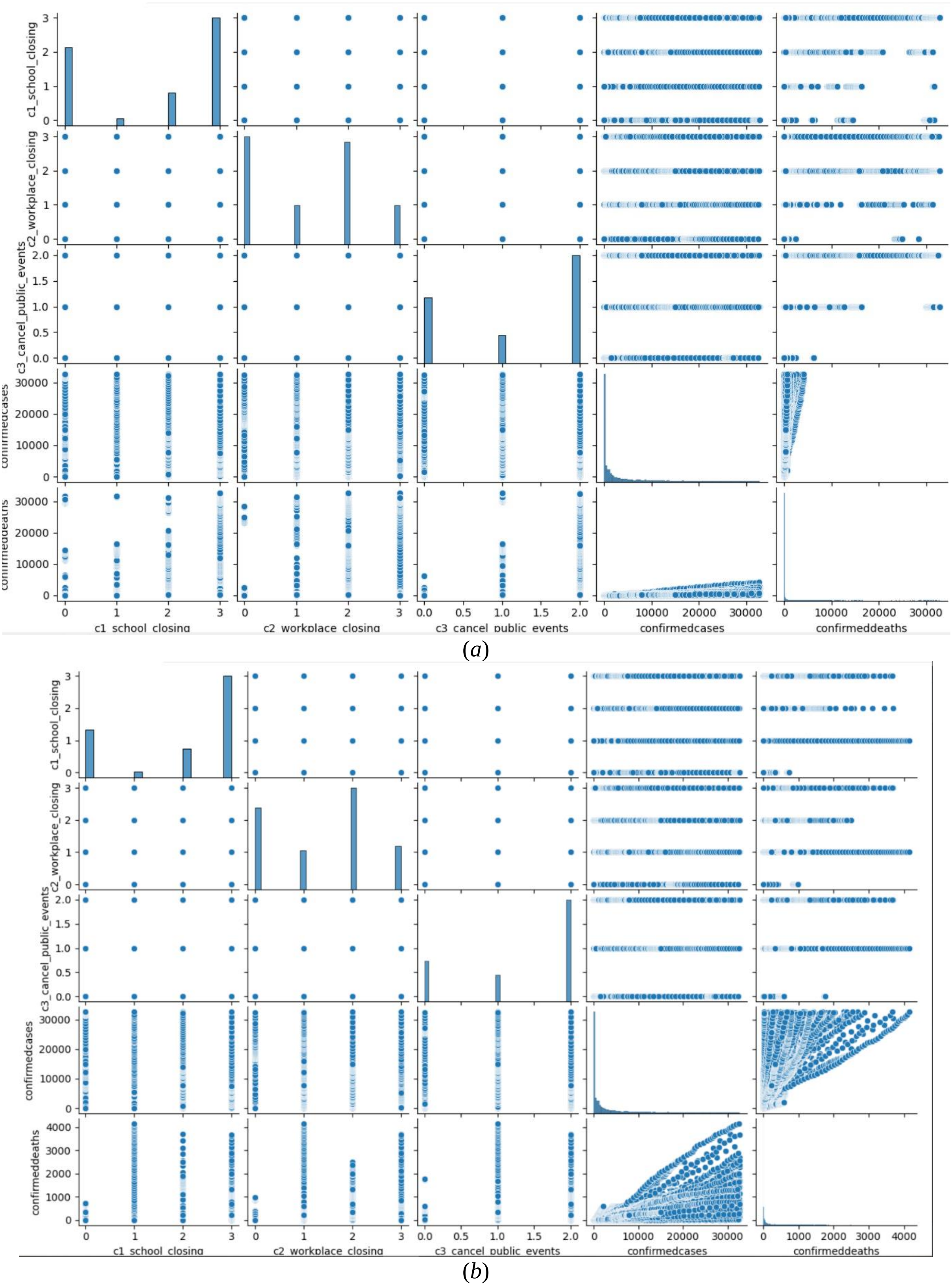

Figure 1: Data distribution before (a) and after (b) preprocessing

The next step is clustering. The *k*-means method is used. The gap-statistics allows to find the appropriative number of clusters.

Cluster centroids allow to find “average” object in each group and to create the regularization rules. The cluster # 3 shows countries under restriction, cluster #1 consists of countries without almost any restriction (see Table 1, Table 2, Table 3).

Table 1

Feature wights for the three clusters obtained

Cluster	School closing	Workplace closing	Cancel public events	Close public transport
1	0.04368254	0.02120635	0.07822222	0.005587302
2	2.89865433	2.26230025	1.94790791	1.109703532
3	1.88113534	1.02194531	1.30058844	0.256905504

Table 2

Feature wights for the three clusters obtained cd.

Cluster	Stay at home req.	Restrictions on internal movement	International travel controls
1	0.0351746	0.01853968	1.137778
2	1.7232443	1.68355761	3.438709
3	0.5112496	0.50931118	2.850190

Table 3

Feature wights for the three clusters obtained cd.

Cluster	Income support	Public information campaigns	Testing policy	Contact tracing
1	0.02933333	0.7869206	0.6651429	0.6673016
2	1.00667578	1.9615749	1.5702271	1.3380467
3	1.17223953	1.9540325	1.8985808	1.6339218

The system of recommendations based on objects distribution in each cluster is depicted in Figure **2**.

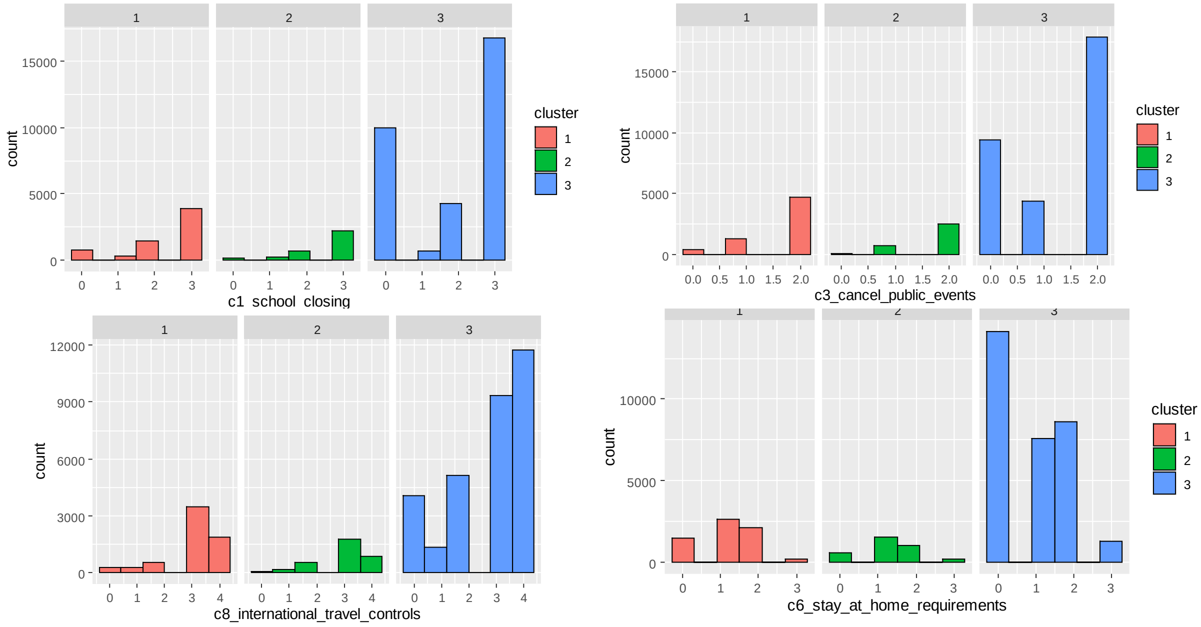

Figure 2: Data distribution among clusters by recommendation based on objects distribution

As it can be seen from the Figure 2 the clusters differ by the most distinct recommendations, which can be summarized as follows:

- Cluster #1: recommended to close schools, and control international travels,
- Cluster #2: recommended to stay at home,
- Cluster #3: recommended to stay at home and cancel public events.

The influence of such clustering on COVID-19 spreading, peak position and duration as well as mortality rate will be the subject of our further study.

An example of countries along with the most frequent cluster number is given in Table 4. The same country can be joined to different clusters in different time slots. So, the clustering by country will not be so unambiguous. That is why time series for separated country can be interesting and will be the subject of the further study. At this stage we used the classification routine on the dataset. The comparison of rule-based, statistical and recurrent neural network methods can also be found at [22].

Table 4
The list of countries (fragment) belonging to different clusters

Country	Cluster
Afghanistan	1
Albania	1
Algeria	2
Andorra	1
Austria	2
Poland	2
Portugal	2
Puerto Rico	1
United Kingdom	3
United States	1
Uruguay	1
Uzbekistan	1
Venezuela	1
Vietnam	1
Yemen	2
Zambia	1
Zimbabwe	1

Three classifiers are analyzed: random forest (500 trees), logistic regression and XGBOOST (tree learning algorithm). The scores of the models are listed in Table 5.

Table 5
The score of the different classifiers

Random forest	Logistic regression	XGBOOST
0.96	0.91	0.89

The rules given below presents the strategy based on decision tree and can be used for strategic planning in the public health system to avoid deaths and severe consequences of the epidemic.

```
if ( confirmedcases <= 11577.5 ) {
  if ( confirmedcases <= 3806.5 ) {     [[confirmeddeaths=20.21931171]]    }
    else {     [[ confirmeddeaths =206.84760845]]    }
 } else {
  if ( confirmedcases <= 17153.5 ) {
    if ( e1_income_support <= 1.5 ) {      [[confirmeddeaths=370.77655172]] }
```

```
     else {
      if ( c8_international_travel_controls <= 2.5 ) {
        if ( c6_stay_at_home_requirements <= 0.5 ) {   [[confirmeddeaths=452.20754717]] }
         else {    [[confirmeddeaths=1410.8028169]]      }
      } else {    [[confirmeddeaths=501.40616622]]     }
     }
   } else {
     if ( e1_income_support <= 1.5 ) {
      if ( confirmedcases <= 24529.0 ) {  [confirmeddeaths= [503.03383459]] }
       else {         [[confirmeddeaths=780.60655738]] }
     } else {
      if ( c8_international_travel_controls <= 2.5 ) {
        if ( c5_close_public_transport <= 0.5 ) {       [[confirmeddeaths=898.76146789]] }
         else {     [[confirmeddeaths=1899.2962963]]          }
       else {
        if ( c6_stay_at_home_requirements <= 0.5 ) {
          if ( c7_flag <= 0.5 ) {
            if ( c7_restrictions_on_internal_movement <= 0.5 ) {          [[confirmeddeaths=1346.375]] }
             else {            [[confirmeddeaths=413.04]]  }
          } else {             [[confirmeddeaths=3256.]] }
        } else {
          if ( confirmedcases <= 29128.0 ) {
            if ( c2_workplace_closing <= 2.5 ) {            [[confirmeddeaths=665.88839286]]  }
             else {
              if ( c5_flag <= 0.5 ) {
                if ( c1_school_closing <= 2.5 ) {              [[confirmeddeaths=2504.]] }
                 else {                [[confirmeddeaths=815.2952381]]  }
              } else {         [[confirmeddeaths=1627.93548387]] }
            }
          } else {
            if ( h2_testing_policy <= 2.5 ) {
              if ( c8_international_travel_controls <= 3.5 ) {              [[confirmeddeaths=1457.]] }
               else {          [[confirmeddeaths=530.2173913]]  }
            } else {             [[confirmeddeaths=708.86764706]]        }
          }
        }
      }
     }
   }
 }
```

5. Conclusions and Future Work

The current COVID-19 pandemic poses serious challenges for data scientists and AI professionals due to limited and incomplete data and the large amount of heterogeneous data, as well as the significant impact of various factors on disease behavior and spread.

This paper is devoted to building the recommendation rules for appropriated state policy for reducing the spread of new Covid-19 cases. The system of these rules is based on novel ensemble of machine learning methods such as regression tree and clustering. The merged data from the Oxford COVID-19 Government Response Tracker and ECDC Covid-19 Cases datasets were used in this study.

The clustering was carried out using the *k*-means method. The gap-statistics allows to find the appropriative number of clusters, and in the case of the study three clusters were selected. The clusters differ by the recommendations and actions made by the correspondent governments. Thus the first cluster countries chose to close schools and to control international travels as main recommendations; the second cluster countries recommended to stay at home, while the major recommendations of governments of countries belonging to the third cluster were to stay at home and to cancel public events.

The same country can be joined to different clusters in different time slots. So, the clustering by country will not be so unambiguous. That is why time series for separated country can be interesting and will be the subject of the further study. Besides, the influence of such clustering on COVID-19 spreading, peak position and duration as well as mortality rate will also be the subject of our further study.

The regression decision tree was built, and the set of rules was extracted from the decision tree and can be used for strategic planning in the public health system.

We believe that the results obtained in this article will contribute to the public good for solving the current global problem. It is also seen that it is necessary to reduce the infection rate sharply and quickly to see a noticeable decrease in the epidemic peak and the death rate.

6. References

[1] COVID-19 coronavirus pandemic. URL: https://www.worldometers.info/coronavirus/

[2] WHO director-general's opening remarks. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020

[3] Zhang, X., Ma, R., Wang, L. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major western countries. Chaos Solitons Fractals 135 (2020), 109829. https://doi.org/10.1016/j.chaos.2020.109829

[4] A. Doanvo, X. Qian, D. Ramjee, H. Piontkivska, A. Desai, M. Majumder. Machine Learning Maps Research Needs in COVID-19 Literature. Patterns (2020), article in press. https://doi.org/10.1016/j.patter.2020.100123

[5] A.W. Salehi, P. Baglat, G. Gupta. Review on machine and deep learning models for the detection and prediction of Coronavirus. Materials Today: Proceedings (2020), article in press. https://doi.org/10.1016/j.matpr.2020.06.245

[6] J.K. Davis, T. Gebrehiwot, M. Worku, W. Awoke, A. Mihretie, D. Nekorchuk, M.C. Wimberly. A genetic algorithm for identifying spatially-varying environmental drivers in a malaria time series model. Environ. Model. Softw. 119 (2019), pp. 275–284. https://doi.org/10.1016/j.envsoft.2019.06.010

[7] M.H.D.M. Ribeiro, R.G. da Silva, N. Fraccanabbia, V.C. Mariani, L.d.S. Coelho. Forecasting epidemiological time series based on decomposition and optimization approaches. In: 14th Brazilian computational intelligence meeting (CBIC), Belém, Brazil, PA (2019), pp. 1–8.

[8] J.M. Scavuzzo, F. Trucco, M. Espinosa, C.B. Tauro, M. Abril, C.M. Scavuzzo, A.C. Frery. Modeling dengue vector population using remotely sensed data and machine learning. Acta Trop. 185 (2018), pp. 167–175. https://doi.org/10.1016/j.actatropica.2018.05.003

[9] Ribeiro, M.H.D.M., da Silva, R.G., Mariani, V.C., dos Santos Coelho, L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons & Fractals 135, (2020), 109853. https://doi.org/10.1016/j.chaos.2020.109853

[10] D. Fanelli, F. Piazza Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals, 134 (2020), p. 109761. https://doi.org/10.1016/j.chaos.2020.109761

[11] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P.Yan, G.Chowell. Real-time forecasts of the COVID-19 epidemic in China from february 5th to february 24th, 2020. Infect. Dis. Model. 5 (2020), pp. 256–263. https://doi.org/10.1016/j.idm.2020.02.002

[12] R. Vaishya, M. Javaid, I.H. Khan, A. Haleem. Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab. Syndrome 14 (4), (2020), pp. 337–339. https://doi.org/10.1016/j.dsx.2020.04.012

[13] Basu, S., Campbell, R.H. Going by the numbers : Learning and modeling COVID-19 disease dynamics. Chaos, Solitons & Fractals 138 (2020), 110140. https://doi.org/10.1016/j.chaos.2020.110140

[14] R.A. Neher, R. Dyrdak, V. Druelle, E.B. Hodcroft, J. Albert. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. Swiss Med Weekly, 150 (1112) (2020). https://doi.org/10.4414/smw.2020.20224

[15] COVID-19 Scenarios. URL: https://neherlab.org/covid19/

[16] Hu Z, Ge Q, Jin L, Xiong M. Artificial intelligence forecasting of COVID-19 in China. arXiv preprint arXiv:200207112 2020. http://arxiv.org/abs/200207112

[17] Bandyopadhyay SK, Dutta S. Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release. medRxiv 2020. https://doi.org/10.1101/2020.03.25.20043505

[18] Z. Yang, Z. Zeng, K. Wang, S.-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, 12 (3) (2020), p. 165. https://doi.org/10.21037/jtd.2020.02.64

[19] S. Tuli, S. Tuli, R. Tuli, S.S. Gill. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things 11 (2020), 100222. https://doi.org/10.1016/j.iot.2020.100222

[20] COVID-19 Data Lake. URL: https://azure.microsoft.com/en-US/services/open-datasets/catalog/covid-19-data-lake/

[21] Boyko, N., Mochurad, L., Stetsiv, I., Kryvenchuk, Yu.: Modeling of the Information System for Processing of a Large Distilled Data for the Investigation of Competitiveness of Enterprises. In: Proc. of the 4th Intl Conf. on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference Lviv, Ukraine, April 23–24, 2020, pp. 964–978. CEUR-WS.org, online CEUR-WS.org/Vol-2604/paper64.pdf

[22] Boyko, N., Mochurad, L., Parpan, U., Basystiuk, O.: Usage of Machine-based Translation Methods for Analyzing Open Data in Legal Cases. In: Proc. of the Intl Workshop on Cyber Hygiene (CybHyg-2019) co-located with 1st International Conference on Cyber Hygiene and Conflict Management in Global Information Networks (CyberConf 2019), Kyiv, Ukraine, November 30, 2019, pp. 328–338. CEUR-WS.org, online CEUR-WS.org/Vol-2654/paper26.pdf