

# The Problem of Analysing the Relationships between Individual Characteristics of Individuals with COVID`19

Nataliia Melnykova<sup>a</sup>, Natalya Shakhovska<sup>a</sup>, Volodymyr Melnykov<sup>b</sup>, Mykola Logoyda<sup>a</sup> and Yulia Peleshchak<sup>a</sup>

<sup>a</sup> Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, Ukraine

<sup>b</sup> Danylo Halytsky Lviv National Medical University, 69 Pekarska str., Lviv, 79010, Ukraine

## Abstract

The paper proposes the development of an approach to modelling the nature of individual morbidity based on the Big Data approach. Analysis of large amounts of data requires the definition of groups of attributes that form functional dependencies. However, in real datasets obtained from different sources, important relationships are defined only for a subset of attribute group values. There are relationships, for example, between previously transmitted diseases and the nature of the disease now - such a relationship is established between subsets of values of different tuples and cannot be found existing methods of searching for hidden data. The authors will call such dependences partial functional dependencies. Accordingly, the level of support for such dependencies is low, which does not allow to use them for further data analysis. At the same time, partial functional dependencies are modified associative rules, but they are executed only for a part of the data and depend on the time factor. The method of finding such dependencies will be based on the modification of the method of associative rules, which allows to reduce the time complexity and to use parallel and distributed mode for calculation.

## Keywords 1

Big Data, associative rules, dependency analysis, Covid`19

## 1. Introduction

It is a well-known fact that artificial intelligence really gives us many opportunities to understand and solve many complex problems in the practical and scientific space. The use of artificial intelligence can be useful primarily in systems that are designed to detect, track and predict disease outbreaks. The better we can track the spread of the virus, the more effective and faster we can fight it.

By analysing press reports, social networking platforms, and government documents, artificial intelligence can learn to detect outbreaks faster. It even turns out that such systems already exist and work, for example, the Canadian system for launching BlueDot [1]. The company's software is designed to protect against the risk of pandemic outbreaks and protects lives by reducing the impact of infectious diseases that threaten human health. The developers claim that their software allows you to convey all the information about the threat of the COVID`19 epidemic within a few hours to the Centres for Disease Control and Prevention or the World Health Organization. The app also helps detect potentially infected people.

The Chinese surveillance system used SenseTime's face recognition technology and software to identify people who may have a fever. The Chinese government has also developed a monitoring system called the Health Code, which uses a variety of data to identify and assess each person's risk based on their behaviour, namely travel history, time spent in hotspots, and potential contact with those who

---

IDDM`2020: 3<sup>rd</sup> International Conference on Informatics & Data-Driven Medicine, November 19–21, 2020, Växjö, Sweden  
EMAIL melnykovanatalia@gmail.com (A. 1); natalya233@gmail.com (A. 2); v.melnikov2013@gmail.com (B. 3); logoyda47@ukr.net (A. 4); yulia.peleshchak@gmail.com (A. 5)  
ORCID: 0000-0002-2114-3436 (A. 1); 0000-0002-6875-8534 (A. 2); 0000-0002-7008-4014 (B. 3) 0000-0001-6476-7305 (A. 4); 0000-0003-1057-6743 (A. 6)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

carries the virus. Citizens are assigned a colour code (red, yellow or green), which they can access through the popular WeChat or Alipay applications for verification [2]. Also, to investigate this problem, the researchers used a convolutional neural network model to detect patients with COVID-19 using CXR images. They used pre-trained ImageNet and trained the model based on open source X-ray imaging Chest (CXR) [3]. Other researchers have used of the LSTM model to predict country-specific risk of COVID-19 virus infection, based on country-specific trends and meteorological data, to predict the likely spread of COVID-19 disease [4].

Artificial intelligence experts have used machine learning techniques to process online activity, news, health organizations' reports and media activities to predict the spread of the outbreak in China. As well as the application of the Bayesian approach to predict the number of deaths in the future, using empirical data [5].

An important role is played by the problem of analysing the relationships between the individual characteristics of the patient and the characteristics of dynamics of changes in his condition during treatment and recovery.

The course of diseases, that caused by infections and viruses (even with known treatment prevention schemes) is influenced by various factors, namely [6,9]:

- variability of strains,
- nature of interaction,
- features of the distribution area: climatic conditions, development of infrastructure and connections, quality of medical care, chronic diseases inherent in this area, political situation, etc.

## 2. The main material

The peculiarity of medical data is their hierarchy and networking. Network data include information on comorbidities, allergic reactions, etc. These are the direct or indirect factor, which determines the nature of disease of the individual. Therefore, to find the hidden dependencies of the data and to determine the nature of the disease of the individual, it is necessary to find not only linear dependencies in the data.

The task of finding dependencies in the data requires the analysis of dependencies between dozens of parameters of the studied process and hundreds of possible sources of influence on this process. Dependencies are nondeterministic, and therefore modelling requires the use of statistical methods for analysing random processes. Often much of the information is hidden from observation or is not monitored. This introduces many difficulties in the process of analysing the collected information.

The modelling process requires data that can be obtained from known statistics, namely:

- the population of the districts of the region;
- the population density;
- the social distance between people;
- the disease duration;
- the probability of disease in human contact; mortality rate;
- the availability of crowded places (supermarkets, churches, pharmacies, markets, construction sites, gyms);
- the percentage of people who carry the disease asymptotically;
- the presence of the procedure of isolation of sick people;
- the ability to move a person from the area to the regional centre and back;
- the incubation period
- and others.

The above factors can negatively affect the conduct, interpretation and generalization of research results, as well as the understanding and interpretation of the phenomenon under study. For stability of result, it is necessary to use ensembles of models, which are easy to parallelize.

The analysis of large amounts of data requires the definition of groups of attributes that form functional dependencies. However, in real data sets obtained from different sources, important relationships are defined only for a subset of attribute group values. For example, between previously transferred disease and the nature of the disease now - such a relationship is established between subsets of values of different tuples and cannot be found by existing hidden methods. We can call such dependences partial functional dependencies. Accordingly, the level of support for such dependencies is low, which does not allow to use them for further data analysis. At the same time, partial functional dependencies are modified associative rules, but they are executed only for a part of the data and depend on the time factor. The method of finding such dependencies will be based on the modification of associative rules. This reduces time complexity and uses parallel and distributed mode for calculation.

As part of the study of the problem is necessary to:

- propose the development of already special methods of forming a training set of data and preprocessing of attributes, taking into account the specifics of the content of medical data and environmental data.
- develop an ensemble of data imputation models on the basis of basic models of various nature as a part of specialized information technology of recovery of the missed data for the automated processing of information.
- simulate different scenarios of influence by the state at the next stage of forecasting the dynamics of infection.

## 2.1. The formal model of the individual

Therefore, for a formal representation of the individual's condition, the task of which is to find a personalized assessment and find solutions to improve it, it is necessary to build a formal model of the object, as a model of production expert system, which is usually used to solve this class of problems.

The knowledge base in accordance with the structural scheme of the system of personalized assessment is the selection of a set of rules  $R$  [7,12,14]:

$$R = \{R_1, \dots, R_j\}, \quad (1)$$

where the production of type

$$R_j = pd_{j1} \wedge pd_{j2} \wedge \dots \wedge pd_{jn} \rightarrow k_m \quad (2)$$

Thus, the production  $R$  is a ratio, the elements of which are indicators of the characteristics of the state of the individual and the assessment of the state of the individual.

$$R: PD \rightarrow VS$$

$$R \subset PD \times VS = \{(a, v): a \in PD \& v \in VS\} \quad (3)$$

where

$$PD = PD(a_n), \\ pd_n \in \{A_t \cup A_{in}\},$$

So

$$PD = \{\{A_{in}\} \cup \{A_t\}\}, \quad (4)$$

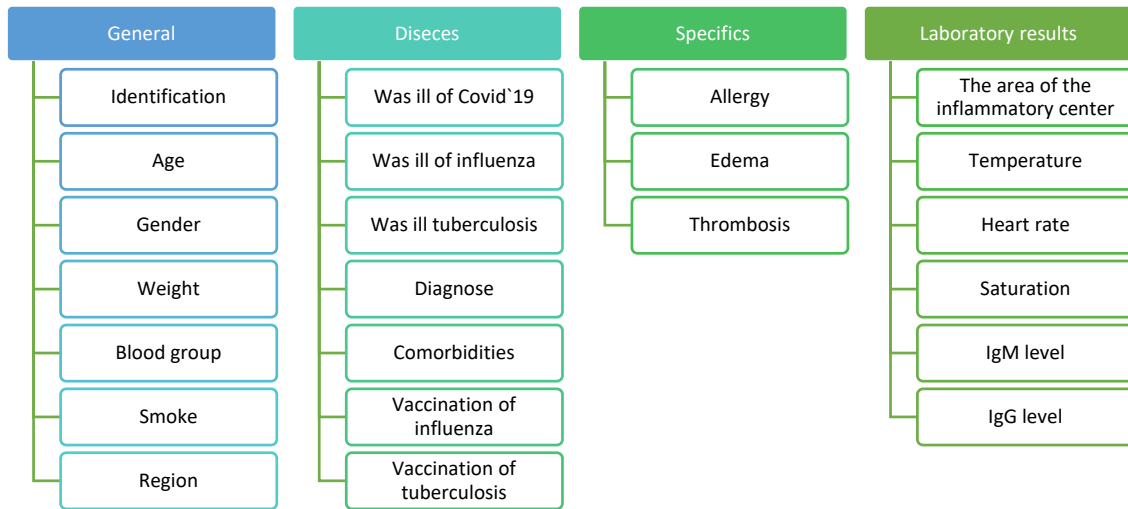
$PD$  – is a set whose elements are the parameters of the state of the individual, namely the elements of sets of time-independent characteristics ( $A_{in}$ ) and time-dependent parameters of the object ( $A_t$ ).

An example of rules is the search for optimal conditions and strategic decisions for the treatment of the individual based on selected characteristics and parameters.

The composition of many informative factors depends on the specific task. In the general case, the dependencies can be divided into linear and nonlinear. In addition, if for linear dependencies informative factors are determined by known methods of correlation and analysis of variance, then for nonlinear dependencies such procedures are often empirical. In our case, the dependencies are predicted as linear [8,10,11].

Among the parameters of the individual to determine the presence of the disease, the data are divided into time-dependent and independent. The parameters must be prepared before the analysis, namely it

is necessary to determine the priority data that will affect the results. The importance of finding relationships between data optimizes the process of determining an individual's resistance to COVID`19 [21].



**Figure 1:** Parameters of the individual to determine the presence of the disease.

## 2.2. The use of associative rules for study dependencies

The basic concepts in the theory of associative rules are subject set and transaction. A thematic set is a non-empty set of elements that can be part of a transaction:

$$I = \{i_1, i_2, \dots, i_k, \dots, i_n\}, \quad (5)$$

where  $i_k$  - are the elements included in the subject sets,  $k = 1..n$ ,  $n$  - is the number of elements of the set I.

The database has a certain set of transactions:

$$T = \{t_1, t_2, \dots, t_i, \dots, t_m\}, \quad (6)$$

where  $t_i$  - is the corresponding transaction,  $m$  - is the total number of transactions.

The concepts of set and associative rule are closely related to another characteristic of an associative rule - trust, which is calculated as the ratio of a set that has both a condition and a consequence (in other words, support for an associative rule) to support a set that has only a condition.

$$Conf(X \rightarrow Y) = Supp(X \rightarrow Y) / Supp(X) = \quad (7)$$

$$\begin{aligned} & (|X(t) \cap Y(t)| / |T|) / (|X(t)| / |T|) = \\ & = |X(t) \cap Y(t)| / |X(t)|, \end{aligned}$$

MinSupp and MinConf minimum support and validity thresholds are used to determine the significance of the rules, which are usually determined by experts based on their own experience:

$$Supp(X \rightarrow Y) \geq MinSupp \quad (8)$$

$$Conf(X \rightarrow Y) \geq MinConf \quad (9)$$

Methods for finding associative rules find all associations that meet the constraints of support and confidence. However, this leads to the need to review a large number of associative rules, which it is desirable to reduce in such a way as to analyse only the most important of them.

Among the main algorithms for generating associative rules are AIS, SETM, Apriori, AprioriTid, AprioriHybrid [8]. The efficiency and feasibility of using each of them is determined by the structure and scope of the data set for which the search for associative rules, as the basis of these methods lies in the different principles of generation and selection of subject sets – candidates [13,15].

The disadvantages of the above algorithms are solved by the Apriori algorithm that proposed by R. Agrawal and R. Srikant. Unlike AIS and SETM, it eliminates the generation and counting of excessive numbers of candidates through the using of the antimonotonic properties and can significantly reduce the set of frequent sets of items and thus reduce the search space for associative rules. The property of diversity states that if the set  $Z$  is not common, then adding a new object  $Y$  to the set  $Z$  does not change its frequency (respectively, if  $Z$  is not a frequent set, then  $ZY$  is not frequent either). Modifications of the classic Apriori algorithm are AprioriTid and AprioriHybrid [16,19,20].

Using the method of a priori implement the search for associative rules. Since the size of modern databases can reach quite large volumes (gigabytes and terabytes), the search for associative rules requires efficient algorithms that are scalable and allow you to find a solution to this problem in a reasonable time [17,18].

Algorithms for recognizing patterns with learning assume the presence of a retrospective, which allows you to build statistical models of dependencies  $x \rightarrow y$ , where  $y \in Y$ ,  $Y$  user actions (answers) are observed or a random variable is modelled  $x \in X$ ,  $X$  - a set of variables (predictors) that are supposed to explain the variability of the variable  $y$ . Most models with a teacher are designed in such a way that they can be written as

$$y = f(x, \beta) + \varepsilon, \quad (10)$$

where  $f$  - a mathematical function selected from some arbitrary family,  $\beta$  - the vector of parameters of this function,  $\varepsilon$  - errors that usually generate unbiased, uncorrelated random processes.

When building a model at fixed sample values,  $y$  minimizes the residuals of some function  $Q(y, \beta)$ . As a result, found  $\beta$ . This is a vector with optimal estimates of the model parameters. By changing the shape of the functions  $f$  and  $Q$ , you can get different models, of which the most efficient model is preferred. This model provides unbiased, accurate, and reliable  $y$  predictions.

### 2.3. The proposed method

Based on the data analysis, information was collected about patients taking into account the parameters of the individual to determine the presence of diseases that are characterized as a priority.

Dataset is collected using google form [https://docs.google.com/forms/d/1o8CMGVZv6BDkw-QIYg2F8VQzqcXxqklomRwXCLOZiC\\_Y](https://docs.google.com/forms/d/1o8CMGVZv6BDkw-QIYg2F8VQzqcXxqklomRwXCLOZiC_Y), is funded by Central European Initiative and is verified by Lviv regional centre Covid`19 resistance. This dataset consists of the following characteristics:

- Age (categorical): 0-15, 16-22, 23-40, 41-65, >66
- Gender (categorical): male, female
- Region (string): Lviv (Ukraine), Chernivtsi (Ukraine), Belarus, Germany, Other
- Do you smoke (Boolean): yes, no
- Have you had COVID`19 (categorical): yes, no, maybe
- IgM level (numerical): [0..0.9) (negative), [0.9..1.1) (indefinite), >=1.1 (positive)
- IgG level (numerical): [0..0.9) (negative), [0.9..1.1) (indefinite), >=1.1 (positive)
- Blood group (numerical)
- Do you vaccinated influenza? (categorical): yes, no, maybe
- Do you vaccinated tuberculosis? (categorical): yes, no, maybe
- Have you had influenza this year? (categorical): yes, no, maybe
- Have you had tuberculosis this year? (categorical): yes, no, maybe

Taken into account 480 responses are presented in dataset.

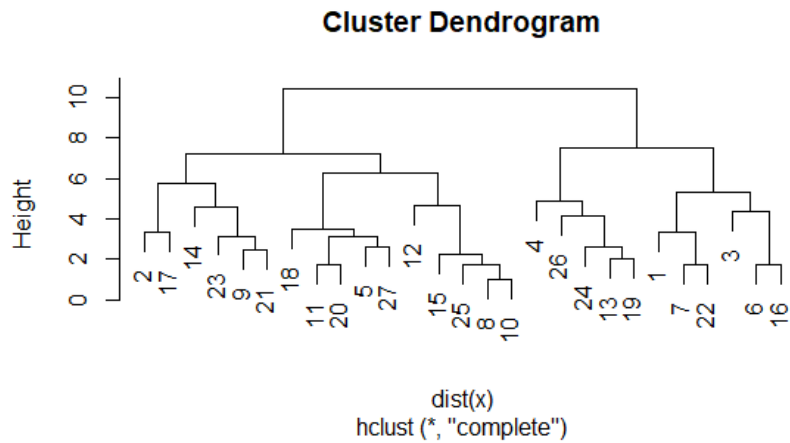
Sample for training, consisting of states of 30 individuals. For data analysis, it was proposed:

- grouping of data by patient ID,
- separation of factors by patient ID,
- division into factors according to the patient's condition,
- separation of factors according to the treatment scheme.

A three-step algorithm was used to recognize the patterns:

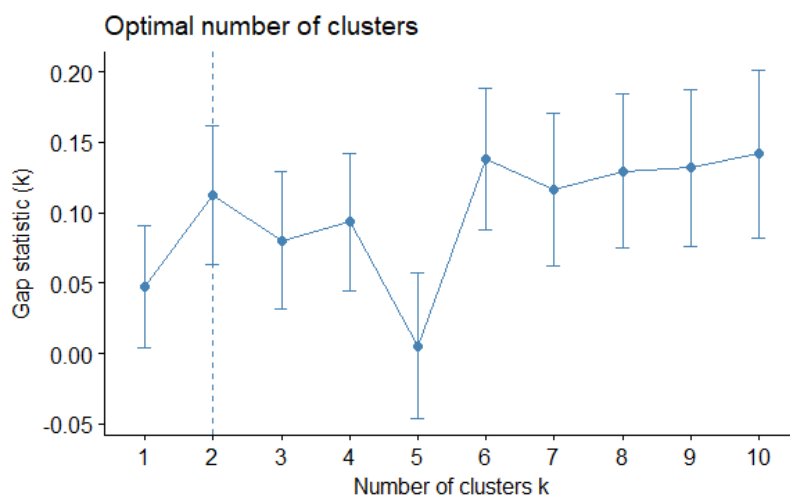
- creating a cluster of patients - to find the behavior of the condition,
- template construction - to find the sequence of state changes,
- next run ConditionType - to predict the next condition of the patient.

The cluster of the studied objects was created by means of *R*, factoextra packages (Fig. 2 shows the results of hierarchical clustering):



**Figure 2:** Dendrogram of hierarchical clustering

In this work we use k-means and k-medoid separation algorithms. First, we found the optimal number of clusters using the Elbow method and gap statistics. Gap statistics can be applied to any clustering method. It compares the total variation of the internal cluster for different values of *k* with their expected values at the null benchmark distribution data (ie distribution without explicit clustering). The reference data set is generated by Monte Carlo simulation of the sampling process [22]. (Fig. 3):



**Figure 3:** The optimal number of clusters

The x-axis ranges from zero trees to the maximum number of trees in which any variable was used for splitting which is in this case equal to 500 and is reached by all variables plotted. Therefore, the

maximum depth in created trees is for vaccinated influenza. The first level in the biggest part of “poor classifiers” is presented by IgG.

For further explore variable importance measures we pass our forest to measure\_importance function and get the following data frame.

	variable	mean_min_depth	no_of_nodes	mse_increase	node_purity_increase	no_of_trees	times_a_root
1	Age	1.511688	3023	0.050938625	11.558790	499	112
2	Blood_group	1.820000	3387	0.013201974	9.690663	500	53
3	Gender	1.723688	1484	0.052533296	6.322656	499	111
4	Had_influenz	1.727688	2225	0.034935393	7.862122	499	92
5	Smoke	2.030000	1507	0.005027793	4.138718	500	79
6	Vaccinated_influenza	2.372752	1896	-0.008977258	3.495168	496	25
7	Vaccinated_tuberculosis	2.164000	2348	0.015751848	5.870700	500	28

We can see that all depicted measures are highly correlated.

After selecting a set of most important variables, we can investigate interactions with respect to them, i.e. splits appearing in maximal subtrees with respect to one of the variables selected. To extract the names of 5 most important variables according to both the mean minimal depth and number of trees in which a variable appeared, we have the following result:

[1] "Age" "IgG" "Blood\_group" "had\_influenz" "IgM"

Naive Bayes shows the density for each features in the dataset (Fig. 4). The accuracy of Naive Bayes is much less than random forest and is equal to 67%.

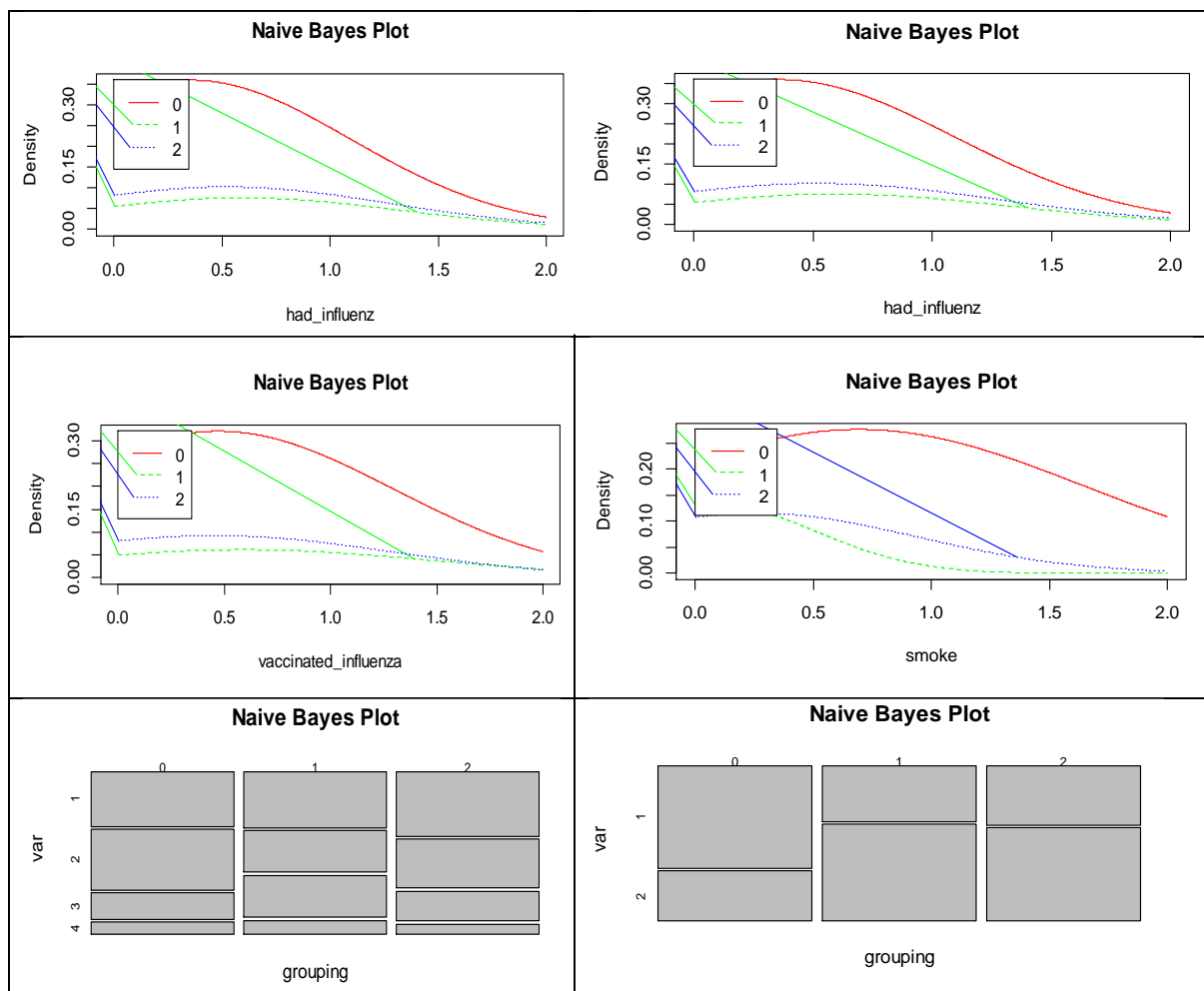


Figure 4: Naive Bayes plot of density

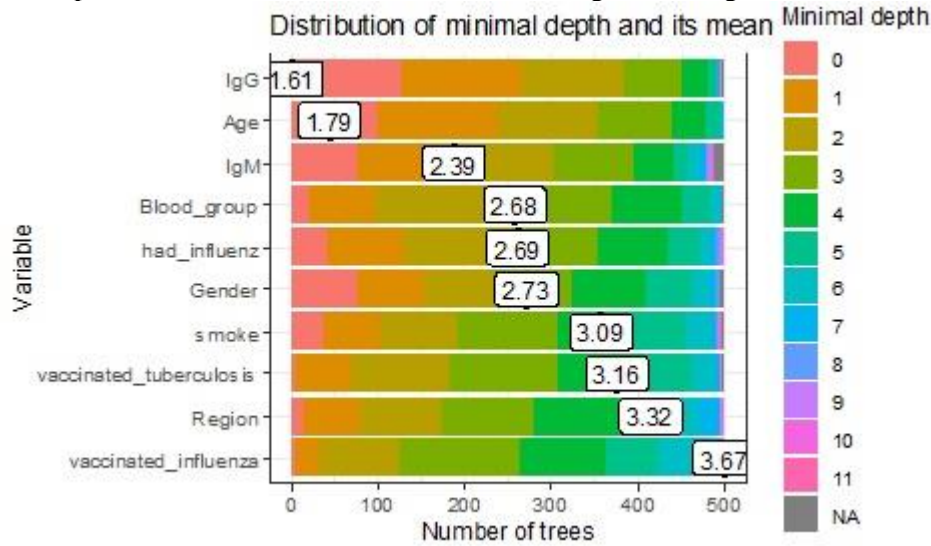
Random forest shows the better results. 500 trees are built.

OOB estimate of error rate: 16.61%

```
Confusion matrix:
0 1 2 class.error
0 153 9 17 0.14525139
1 8 88 12 0.18518518
2 1 3 20 0.16666666
```

The biggest error is for class 1 (COVID'19 - yes). It can be explained by differences in IgG and IgM representation (data scatter is between 0.00 and 18.00) in different countries.

The minimal depth values for all trees in a random forest is given on Fig. 5.



**Figure 5:** Distribution of minimal depth of developed trees

Building a template

There are objective and subjective measures of compliance with the associative rule [7]. The tasks are the above support and confidence. Subjective measures of significance are the elevator and levers. The rise is determined by the ratio of the preservation of the associative rule to the state of product support and the effect separately:

$$Lift(X \rightarrow Y) = \frac{Supp(X \rightarrow Y)}{Supp(X) * Supp(Y)} \quad (11)$$

Lift is a so-called generalized measure of connection between two subject sets. Its meaning can be interpreted as follows:

$$\text{if } Lift(X \rightarrow Y) = 1,$$

$$\text{then } Supp(X \rightarrow Y) = Supp(X) * Supp(Y), \quad (12)$$

that is, the state and the consequence do not depend on each other;

$$\text{if } Lift(X \rightarrow Y) > 1,$$

$$\text{then } Supp(X \rightarrow Y) > Supp(X) * Supp(Y), \quad (13)$$

that is, the effect is positively dependent on the condition;

$$\text{if } Lift(X \rightarrow Y) < 1,$$

$$\text{then } Supp(X \rightarrow Y) < Supp(X) * Supp(Y), \quad (14)$$



that is, the effect is negatively dependent on the condition.

An algorithm for extracting associative rules is proposed. For the relationship with the scheme

$$R = \{A_i, \overline{dom(A_i)}\}, i = \overline{1, m}, \quad (15)$$

allows you to find statistically significant rules that reflect the dependence of the attribute  $A_m$  on the attributes  $A_1, A_2, \dots, A_{m-1}$ , ie the dependence of the species.  $A_1, A_2, \dots, A_{m-1} \rightarrow A_m$

The Kullback-Leibler information index is used as a measure of statistical significance. The algorithm allows you to search only for dependencies defined by a common set of input data; in addition, it has a high computational complexity if there are many classification rules.

### 3. Conclusion

An approach to modelling the nature of individual morbidity based on the Big Data approach is proposed. Analysis of large amounts of data requires the definition of groups of attributes that form functional dependencies.

Explained on real data sets obtained from different sources, important relationships are defined only for a subset of the values of the attribute group.

Dependencies with partial functional dependencies have a low level of support, which does not allow their use for further data analysis, and partial functional dependencies are modified associative rules, but they are executed only for part of the data and depend on the time factor.

A method of finding dependences is proposed, which is based on the modification of the method of associative rules, which allows to reduce the complexity of time and use parallel and distributed mode for calculation.

### 4. Funding

Funding of article by grant contribution of the Central European Initiative in according project STOP COVID 19, Ref. No. 305.2825-20, CEI.

### 5. References

- [1] <https://root-nation.com/ua/articles-ua/tech-ua/ua-shtuchnij-intelekt-covid-19/>
- [2] <https://qz.com/1803737/chinas-facial-recognition-tech-can-crack-masked-faces-amid-coronavirus/>
- [3] Perumal, Varalakshmi, et al. "Detection of COVID-19 Using CXR and CT Images Using Transfer Learning and Haralick Features." *Applied Intelligence*, Aug. 2020. DOI.org (Crossref), doi:10.1007/s10489-020-01831-z.
- [4] Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864. <https://doi.org/10.1016/j.chaos.2020.109864>
- [5] Bayes, C., Rosas, V. S. y., and Valdivieso, L., "Modelling death rates due to COVID-19: A Bayesian approach", arXiv e-prints, 2020.
- [6] Some Methods for classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297 .
- [7] Shakhovska, Natalya & Kaminsky, R. & Zasoba, E. & Tsiutsiura, M.. (2018). Association rules mining in big data. *International Journal of Computing*. 17. pp. 25-32.
- [8] The personalized approach to the processing and analysis of patients' medical data. CEUR Workshop Proceedings. – 2018. – Vol. 2255:Proceedings of the 1st International workshop on informatics & Data-driven medicine (IDDM 2018) Lviv, Ukraine, November 28–30, 2018., 103-112

- [9] Shakhovska, N., Fedushko, S., Greguš ml., M., Melnykova, N., Shvorob, I., & Syerov, Y. (2019b). Big Data analysis in development of personalized medical system. *Procedia Computer Science*, 160, 229–234. <https://doi.org/10.1016/j.procs.2019.09.461>
- [10] Shakhovska, N., Fedushko, S., Greguš ml., M., Melnykova, N., Shvorob, I., & Syerov, Y. (2019b). Big Data analysis in development of personalized medical system. *Procedia Computer Science*, 160, 229–234. <https://doi.org/10.1016/j.procs.2019.09.461>
- [11] Nataliya Boyko, Lesia Mochurad, Iryna Stetsiv, Yurii Kryvenchuk : Modeling of the Information System for Processing of a Large Distilled Data for the Investigation of Competitiveness of Enterprises // Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference Lviv, Ukraine, April 23-24, 2020. pp. 964-978
- [12] Nataliya Boyko , Lesia Mochurad, Iryna Andrusiak, Yurii Drevnytskyi : Organizational and Legal Aspects of Managing the Process of Recognition of Objects in the Image // Proceedings of the International 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. - 571-59
- [13] Gaudart J, Ghassani M, Mintsá J, Waku J, Rachdi M, Doumbo OK, Demongeot J (2010) Demographic and spatial factors as causes of an epidemic spread, the copule approach: application to the retro-prediction of the black death epidemic of 1346. In: 2010 IEEE 24th International conference on advanced information networking and applications workshops (pp 751–758). IEEE.
- [14] .Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*.(2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
- [15] Pham QV, Nguyen DC, Hwang WJ, Pathirana PN. Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. Preprints. (2020) 2020:2020040383. doi: 10.20944/preprints202004.0383.v1
- [16] Mochurad L., Solomiia A. (2020) Optimizing the Computational Modeling of Modern Electronic Optical Systems. In: Lytvynenko V., Babichev S., Wójcik W., Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) *Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. Advances in Intelligent Systems and Computing*, vol 1020. Springer, Cham. pp 597-608. doi: 10.1007/978-3-030-26474-1\_41 ([https://doi.org/10.1007/978-3-030-26474-1\\_41](https://doi.org/10.1007/978-3-030-26474-1_41))
- [17] Mochurad L., Shakhovska K., Montenegro S. (2020) Parallel Solving of Fredholm Integral Equations of the First Kind by Tikhonov Regularization Method Using OpenMP Technology. In: Shakhovska N., Medykovskyy M. (eds) *Advances in Intelligent Systems and Computing IV. CCSIT 2019. Advances in Intelligent Systems and Computing*, vol 1080. Springer, Cham. – pp. 25-35. DOI: 10.1007/978-3-030-33695-0\_3, 11 p. [https://doi.org/10.1007/978-3-030-33695-0\\_3](https://doi.org/10.1007/978-3-030-33695-0_3).
- [18] D. Peleshko, Y. Ivanov, B. Sharov, I. Izonin and Y. Borzov, "Design and implementation of visitors queue density analysis and registration method for retail videosurveillance purposes," 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, 2016, pp. 159-162, doi: 10.1109/DSMP.2016.7583531.
- [19] Sujatha R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic inIndia. *StochEnvironResRiskAssess*.(2020) 34:959–72. doi: 10.1007/s00477-020-01827-8
- [20] Khamparia A, Gupta D, de Albuquerque VHC, Sangaiah AK, Jhaveri RH. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *J Supercomput*. (2020) 76:1–19. doi: 10.1007/s11227-020-03159-4
- [21] Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR. Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*. (2020) 8:91916–23. doi: 10.1109/ACCESS.2020.2994762
- [22] Sakarkar G, Pillai S, Rao CV, PeshkarA, Malewar S. Comparative study of ambient air quality prediction system using machine learning to predict air quality in smart city. In *Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019)*, NITTTR Chandigarh, India. Singapore: Springer (2020). p. 175–82. doi: 10.1007/978-981-15-3020-3\_16