Baltic J. Modern Computing, Vol. 11 (2023), No. 1, 114-133
https://doi.org/10.22364/bjmc.2023.11.1.07

Methodology for Mathematical Determining Key Performance Indicators of Socioeconomic Processes

Girts KARNITIS[1], Janis BICEVSKIS[1], Maris PUKIS[1],
Ugis SARMA[2], Stanislavs GENDELIS[1], Andris EIHMANIS[1],
Andris VIRTMANIS[2], Edvins KARNITIS[1]

[1] University of Latvia, Raina Blvd. 19, LV-1586, Riga, Latvia
[2] Riga Technical University, Meza Street 1 k1, LV-1048, Riga, Latvia

Girts.Karnitis@lu.lv, Janis.Bicevskis@lu.lv, maris@lps.lv,
Ugis.Sarma@latvenergo.lv, Stanislavs.Gendelis@lu.lv,
Andris.Eihmanis@lu.lv, Andris.Virtmanis@rtu.lv
Edvins.Karnitis@lu.lv

[0000-0002-7563-6383], [0000-0001-5298-9859], [0000-0001-7000-2851],
[0000-0003-4654-4647], [0000-0003-3699-2526], [0000-0003-2176-4010],
[0000-0003-1114-1012] [0000-0001-6784-7327]

Abstract. Most socioeconomic processes (Processes) develop by performing various process-driving actions. A large amount of data is usually generated at the same time; they form specific indicators that are more or less distinctive for the Process. No Process management can really perform all the relevant actions to achieve progress of the whole set of indicators. Hence, prioritization of the action lines and determination of the key performance indicators (KPIs) has become an essential factor. Unfortunately, KPIs and their weighting are still largely subjectively defined and there is a lack of qualitative and quantitative justifications for choices. The article describes the universal methodology developed for objective mathematical computation of KPIs of the Processes and their weighting. By means of the regression analysis algorithms for statistically significant KPIs are computed and mathematical expression has been obtained showing the impact of each selected KPI on the Process development. The methodology has been tested in several Processes, achieving convincing results; applying it to variety of Processes requires mediocre programming skills only. Process management can put the methodology into practice to monitor the achieved development level of the Process in statics and dynamics, to observe progress and deficiencies in separate aspects, to take these into account when making the sustainable planning and strategic decisions.

Keywords: process management, key performance indicators, regression analysis.

This paper is an extended version of paper (Karnitis et al., 2022).

1. Introduction. More data is better, but is it all significant?

1.1 Sharply growing amount of available data – a natural tendency

The ongoing digital transformation of a wide range of socioeconomic processes (Processes), together with the direct benefits, also provides an increase in the availability of various generated data which is collected in public and private data bases. Advanced data gathering technologies and cheap data storage are strong enablers of the data collection tendency. This phenomenon can be expected to continue to expand in the future.

There are two kinds of new data. Part of them only increases the amount of data in the existing data series, which are the basis for indicators that describe various aspects of the respective Process. However, this quantitative change sometimes can generate a qualitative effect; the time series data is a convincing example, especially when advanced data mining algorithms are used ((Wauchope et al, 2021), (Warrier et al., 2022), (Li, 2021)).

Another part of new data creates new data series (indicators) that can qualitatively improve the reflection of the Process and its progress; they emerge or lose their relevance as the Process evolves. Comparing the sets of indicators forming Digital Economy and Society Index (DESI) (WEB, a) and (WEB, b) and European Innovation Scoreboard (EIS) (WEB, c) in 2016 and 2022, it can be seen that the indicators, which have lost their significance as a result of sociotechnical development, have been replaced by more advanced ones. Thus, the indicators reflecting the effect of digital transformation and green course on innovation are added to the set of EIS indicators; the innovation performance currently is assessed against 32 indicators, while in 2016 – only against 25 indicators (for more details on the evolution of DESI see Sections 3.2.1 and 3.3).

It would be natural to believe that the amount of data available and used can improve the quality of and trust in the process analysis (description, assessment, forecast, etc.). Academic discussions demonstrate benefits generated by the use of large amounts of data (Philips-Wren, 2021). Developing new indicators is proposed with a view to improve the assessment of Processes in different sectors: economy ((Li, 2020), (Gabbi, et al., 2021), (Eggenberger et al., 2018), ecology (Zhang, et al., 2020), social sectors (Onel et al., 2019), etc. The purpose of the proposals is treated quite differently, but in practice it is relatively the same: to improve the implementation and management of the corresponding process and to increase its efficiency.

A large number of indicators is used to create various national and municipal scale indicator sets and composite indices. Hence, the Global Competitiveness Index (GCI) is composed of 103 different national-size indicators (WEB, d); Global Innovation Index (WEB, e) consists of 80 indicators, while 15 of them are composite indices, which in turn are made up of many indicators; 141 indicators are approved by Saeima (the parliament of Latvia) to evaluate implementation of the National Development Plan of Latvia for 2021-2027 (WEB, f).

Data processing companies are demonstrating the potential advantages of widespread use of large data massifs, with a particular focus on business processes. “The ability to gather data from disparate sources and analyze it for strategic decision-making is the key to the gold mine” (Polakoff, 2020). However, far not all companies are blindly following these recommendations, apparently carefully considering the pay-back of investment required. According to Eurostat, only 13% of EU enterprises without financial sector (10+ employees) analysed big data internally in 2020 (for comparison – 36% of EU enterprises had one of the ERP software packages in 2020). To accelerate the digital transformation of the industrial ecosystems EC propose that 75% of all European enterprises (including SMEs) would take up big data by 2030 (WEB, g).

Sometimes there is even an over-reliance on the amount of data; this is well illustrated by the public confession of the Google's Chief Scientist Peter Norvig: "We don't have better algorithms than anyone else; we just have more data" (WEB, h).

1.2 Determination of KPIs – a practical necessity

Clearly, in many cases national, sectoral or local management cannot perform all the activities to achieve progress in all the indicators, due to the insufficient capacity and inability to invest simultaneously in all action lines. The same relates to businesses; far not all companies are blindly following these recommendations, apparently carefully considering the pay-back of investment required.

Not without practical reason there is published information in recent years on studies, which shows that sometimes usage of too much data is not purposeful, that it leads to statistics exaggeration – overfitting; this may be direct result of high data availability. "The answer to the question of how much data we really need depends on the use case, as well as who is asking the question" (Karnitis et al., 2021). Not all data series are statistically significant; they could be dropped, but which ones? "Knowing more details about each example doesn't necessarily help. It might even hurt, if the extra details are irrelevant to what you're trying to learn" (Sarma et al., 2019). And for a specific purpose: "The results reveal that, while data variety and velocity positively enhance firm innovation performance, data volume has no significant impact" (Zuters et al., 2016).

Simplicity of the composite index, use of limited number of input indicators are attractive characteristics of the Human Development Index (HDI), which is globally accepted at the various levels; it consists of four indicators only (Sarma et al., 2019). "The HDI's simplicity, coupled with the transparency assured by the utilization of data published by international organizations, has been one of the main drivers behind the success of the HDI in the past twenty years". The aspect of simplicity should not be underestimated: this would be one of the reasons why Google search engine found more than 11 million items on the query "Human Development Index", and only 460 thousand items on the query "Global Competitiveness Index"; Science Direct database provided 7911 and 595 scientific publications respectively.

In order to improve each of the indicators included, the national and / or local government must take various coordinated activities that form several action lines. It is absolutely clear that no government can really grasp and perform all of them to achieve progress, even in part, due both the insufficient capacity to cover the whole very wide spectrum and inability to invest immediately and simultaneously in all action lines. The same relates to businesses. The prioritization of the indicators (and hence, the related action lines) has become an essential, sometimes – even critical factor. This encourages analysts to create methods that determine which indicators from the large data set are decisive Process drivers, the key performance indicators (KPIs). Discussion papers are published (Goldstein, 2022) and (Trerotola et al., 2021) to reflect this issue, presenting the advantages, as well as the disadvantages of large data sets (WEB, a). The larger the number of available data series, the more likely it is that the priority data series (indicators), which best reflect the regularities of the process. Therefore, development of methodologies for selecting KPIs from the total available set as well as their weighting in the composed indices is relatively active. But the identification of KPIs also means the identification of related action lines that are prior for the progress. Prior actions are drivers of the progress, while the indicator measures the progress made.

The first step in selecting a KPI is defining a target (e.g., growth, performance, productivity, sustainability, even creditworthy), followed by creating an appropriate massif of candidate KPIs that characterize the target variable from various aspects. Statistics, surveys, interviews, literature review, research studies, experience are used. Various procedures for selection of KPIs are proposed in a variety of sectors, including economy ((Li et al., 2018), (Tesic et al, 2018), (Kaur et al., 2021)), social services ((Jiang et al., 2020), (You ,2016)), environment (Pakzad et al.,2017) etc.

The methodologies used to select KPIs and to determine their weighting can be grouped into several clusters. It is still popular to rely on the experts' subjective choices in the KPI selection procedure ((Pakzad et al., 2017), (Zacepins et al., 2019), (Kibira et al., 2018), (Cakula, 2020)). In a number of cases, experts directly subjectively carry out the weighting of the selected KPIs too. In other cases, various mathematical methods are additionally used for processing expert assessments, for example, fuzzy analytical hierarchical process (Kaur, 2021), decision making method DEMATEL (Jiang et al., 2020), qualitative scoring method and analytical hierarchical process (Kubiszewski et al., 2021). The results of these calculations, of course, maintain the subjectivity of the experts' assessments.

However, some methodologies, which aim at obtaining as objective as possible selection of KPIs, have also been proposed, involving mathematical tools. Weighting calculations can also be performed at the same time. There are applied data envelop and correlative analyses (Tesic et al., 2018), ontologies (Roldan-Garcia et al., 2022), machine learning algorithms (WEB, i), regression analysis (You, 2016), (Kubiszewski et al., 2021). Unfortunately, the methodologies proposing the application of mathematical tools for selection of KPIs are not opened in referred publications, which make their re-apply for similar tasks practically impossible.

The aim of this study is development of universal transparent methodology for objective mathematical computation of KPIs of the socioeconomic processes and determining their weighting. Objective computation of KPIs (resp., decisive independent input variables) from the set of Process indicators is important for the creation of high-quality Process models, enabling the correct reflection of the impact of key drivers in the model (Karnitis et al., 2022). The actual KPIs are the basis for a model that is accurate as possible and at the same time simple and understandable to its users – non-IT professionals. Furthermore, a transparent KPI computing methodology is one of the key aspects for trust in the model. Therefore, only the application of mathematical tools for determination of KPIs has been considered acceptable.

Selection of the appropriate mathematical tool is described in Section 2, while the Section 3 is devoted to the methodology itself. Section 4 discusses the results obtained and provides conclusions.

2. Regression analysis: a suitable mathematical tool

To achieve the stated aim, a working hypothesis was put forward: let us consider that any Process contains k static units of observation, at which the achieved level of the

Process development is indicated by the characteristic headline indicator (Yk), the value of which depends on values of n KPIs (p1k, p2k,….pnk). Then the Process as a whole can be described by the mathematical model of the Process, where the achieved level of the Process development is described by target/dependent variable (Y) that is a multi-parameter function (f) from the set of KPIs (p1, p2,….pn):

$$Y = f\,(p1, p2, \ldots, pn) \qquad (1)$$

Changes in the value of (Y) reflect the progress achieved. Each KPI shows pro-gress in some aspect of the Process that has been achieved due to the performance of relevant action lines. Thus, the identification of KPIs will also mean the identification of related action lines, which performance is a priority for the Process development, whereas determination of KPIs weighting will show the impact of a specific KPI on the Process.

Today, no direct theoretical calculation is possible for the function (f) and the subsequent measurement of progress in the Process. However, there are several data mining technologies that are suitable for studying cause and effect relationships between x input/independent variables (i1, i2, …, ix) and target variable (Y) without exploring the internal aspects of the Process (black box principle).

Various data mining procedures could be used to simulate socioeconomic processes. The adaptation of the data processing methods for the selection of KPIs is an innovative approach that provides significant benefits, which are shown in the following sections.

The regression analysis was chosen as the most preferable mathematical tool. It is directly focused on the revealing causalities between several independent variables and the dependent variable. Most of regression analysis algorithms do not require normalization of indicators' data; although normalization is widespread and is performed by maintaining the ratio of data point values for a particular indicator, the choice of min/max values affects the inter-indicator comparison. The obtained modelling result is a decoded mathematical expression (1) that shows weighting of each independent variable. In addition it indicates p-value – probability that null hypothesis is taking place and no statistical causality exists between specific independent indicator and target variable; the lower this probability, the higher the statistical significance of the indicator. The result is understandable and convincing for non-IT professionals too. The procedures of analysis are relatively easy to apply.

The authors have modelled several Processes over the years: regularities of EU economies (Karnitis, 2017), the impact of digitalization on economic growth (Karnitis et al., 2018), priority actions for urban sustainable development (Karnitis et al., 2021), and the efficiency of district heating networks (Sarma et al., 2019). For the control of stability and sustainability, a computation of KPIs for the digitalization process was also performed, using highly modified in 2021 set of indicators for recalculated DESI 2019-2020 (for greater detail see Sections 3.2.1 and 3.3).

By means of regression analysis algorithms we have disclosed and extracted the most significant indicators, i.e., the KPIs, reducing a large number of potential Process drivers ($n \ll x$). The practice of applying regression analysis has proven its suitability for the analysis of the status quo and for revealing causal relationships between the independent and dependent variables, as well as for forecasting.

Therefore, the creation of a building thermal efficiency model has been initiated with the aim to determine and select KPIs from the large set of indicators that is currently used for laborious complicated calculations. Whereas up to 75% of the EU around 120 million building stock remains energy inefficient, there is a need for easy-to-use methodology and an appropriate innovative tool, which would allow rapid and at the same time sufficiently accurate and objective analysing at the pre-design stage the potential and benefits of the various complex renovation projects' variants (on the first results see in Sections 3.1.1 and 4).

Specific assumptions to be considered when using regression analysis:

- The model cannot be an abstract representation of the data scope. Whether and why a relationship between two variables has a causal interpretation or why an existing relationship has the power to predict new content should be carefully investigated.
- Not always it is possible to carry out the modeling and select KPIs using standard procedures. Usually, they need to be individually innovatively supplemented and/or modified for a specific Process (see examples in the Section 3). In this aspect, the main requirement is not a high qualification of the software engineer; mutually interested close cooperation of the Process expert, data analyst and programmer is the basis for success when processing observational data.

Obtaining a qualitative result of the modelling requires:

- Full data set (values of independent variables and target variable) in all observation points.
- Similar impact of external factors on all observation points.

Applying mutually absolutely independent modelling tools very different mathematical expressions of the Process model can be obtained, but they usually give very comparable qualitative results, as shown by the usage of different modelling procedures for the same task (Zuters et al., 2016), (Sarma et al., 2019). Thus, Fig.1 shows an excellent coincidence of both the nonlinear regression and neural network models of heat transmission costs.

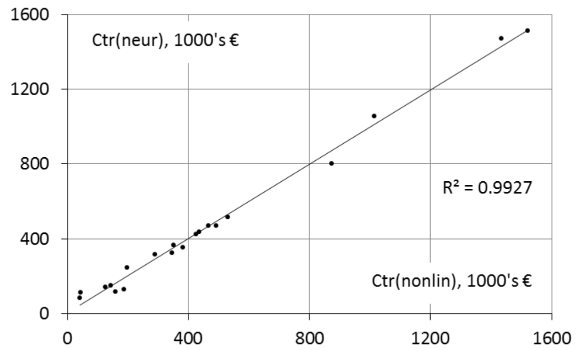

Figure 1. The mutual coincidence of the neuron network model Ctr(neur) and the nonlinear regression model Ctr(nonlin) of heat transmission costs.

3. Methodology: sequential steps

The methodology consists of 5 sequential steps (Fig. 2). The single target (dependent) variable (Y) should be determined in the first step of the modelling procedure. Indicators (independent variables) that describe the target variable as comprehensively as possible from different aspects are selected in the step 2. Computing of KPIs is taking place in the third step by the multi-stage linear regression procedures. In the step 4, the KPIs weights are specified by adding nonlinearity at the level of KPIs and/or the level of the mathematical expression of the model. Varied application-related tests can be applied in the step 5 to check stability and sustainability of the models.

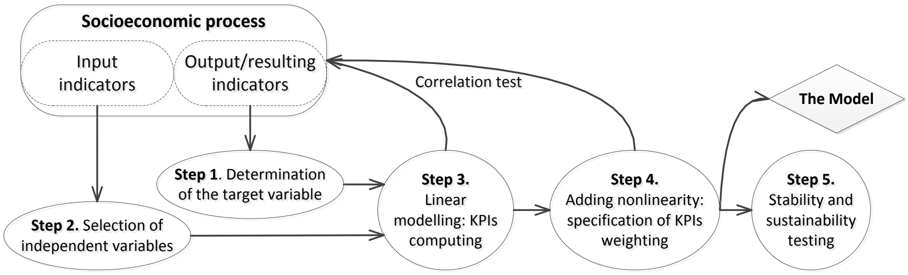

Figure 2. Design of the methodology.

3.1. Selection of variables for modeling

Please note that, when determining the variables for modeling, it is necessary to test the existence of causalities between the target variable and each independent variable. If some correlation is very weak, the particular combination of variables is not suitable for building the model. If, when evaluating the Process, the functional impact of the relevant indicator on the Process is visible, it is worth looking for ways to modify the variable, thus strengthening the existing weak causality.

3.1.1. Determination of the target variable

The most complete and comprehensive headline indicator of the Process is usually used as the target variable to reflect a progress toward the set strategic goal. Result-oriented indicators are preferable, in addition to describing the development process, they characterize the achieved result. Determining the target variable, which is an indicator that characterizes the process in general, is a critical step toward obtaining a reliable result. If the target variable is determined incorrectly (inaccurately), KPIs will also be selected incorrectly.

So, if the Process relates to the national economy, Gross Domestic Product (GDP) or its derivatives can be used. Similarly, the level of the socioeconomic development is generally described by the Human Development Index (HDI) (WEB, d). Both indicators are accepted by experts and politicians. Success of the business activities could be characterized by the profit obtained, performance of the health care system – by treatable mortality, level of education – by years of schooling, efficiency of services at a defined quality/performance – by price (tariff), etc.

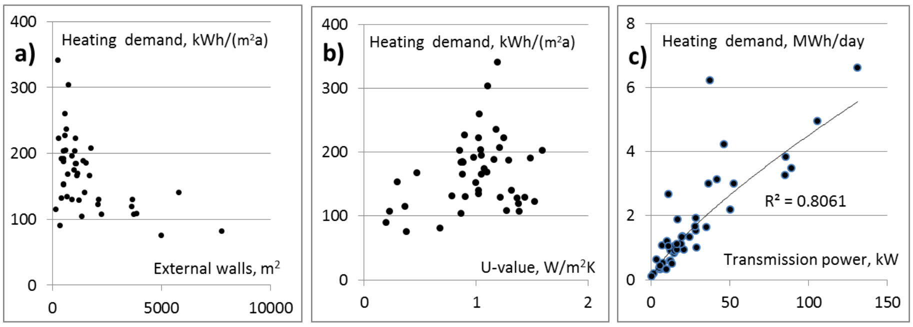

Figure 3. Achievement of the causality by development of the secondary indicators: heating demand (kWh/m^{2}a) dependence on area of external walls (a) and on U-value (W/m^{2}K) (b), causality among secondary indicators (c).

Sometimes it is necessary to modify a widely used target indicator. E.g., the ther-mal performance of a building is usually expressed as the energy use intensity – heating demand per square meter of heated area per year (kWh/m2a). However, the heated area correlates weakly with areas of the building envelope components (external walls, windows, etc.) (Fig. 3a) and their thermal transmittances (U-values, expressed in W/m2K) (Fig. 3b), as well as with other physical and thermal parameters; such regularities are not usable for modelling. For our task we modified the standard indicator developing the total heating demand per heating day (MWh/day). Such a secondary indicator directly reflects the heat transmission process in buildings, forming the necessary causal relationship. Later the modelling result would be converted to a square meter of heated area and to a year. This transformation proved necessary but not sufficient; it turned out that the independent variables also need to be transformed (see Section 3.1.2).

If the corresponding comprehensive indicator is not available, in individual cases it is necessary and possible to create one. For instance, there was no indicator to show the sustainable development (SD) level achieved by countries according to the UN globally accepted SD paradigm – integrity and balance of the economic, social and environmental dimensions. Therefore the appropriate indicator – an Advanced Human Development Index (AHDI) was created to use it as a target variable in modelling the SD process (Karnitis et al., 2021), (Fig. 4).

A purposeful detailed investigation of the cohesion of SD vision and past experi-ence in usage of the HDI for measurement of human development in pre-SD years was carried out, as well as examination of existing proposals A targeted analysis to this end opened several aspects that were considerably taken into account: heredity and simplicity of the target variable, authority of the HDI calculation methodology, the need to use an already accepted comprehensive headline environment indicator. AHDI was created by supplementing HDI with the comprehensive Environmental Performance Index (EPI), which provides an incomparably wider coverage of the environmental aspects in comparison with that of the other options. HDI calculation methodology was precisely adapted to obtain the mathematical expression of the AHDI and to use it as target variable for modelling.

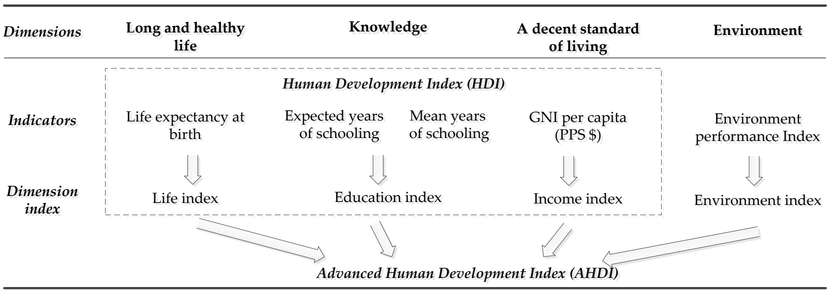

Figure 4. Design of the AHDI.

3.1.2 Selection of independent variables

Sometimes the set of independent variables (indicators) that characterize the Process is already defined (e.g., (Karnitis et al., 2018)), at other times one's own has to be created. In the latter case to create a good model it is important that the selected indicators describe the target variable (Y) as comprehensively as possible, from different aspects. So, for many-sided reflection of the growth of EU economies, and their gradual transformation on innovation-driven growth path, the indicator set of the EIS was supplemented by DESI and energy productivity indices (Karnitis G. et al., 2017). The set of the sustainable development goals (SDGs) have detailed the UN's understanding on the urban SD format, therefore, selected independent indicators (i1, i2, …, ix) should associate with one of the 13 SDGs related to city-level performance (Karnitis et al., 2021); 49 selected indicators were grouped in 13 separate groups related to one of the SDGs. The existence and availability of data sometimes is a constraint on the set formation (this was the case when modelling heat transmission (Sarma et al., 2019).

Also, when selecting the independent variables, it may happen that the traditional indicators have to be modified to improve the causality with target variable. Continuing case of modelling thermal efficiency of buildings, it should be noted that the causal relations of the areas of building envelope's components with the total daily heating demand were found to be sufficient for modelling, while those of the U-values were too weak. Thus, the area of the component (e.g., external walls) was put together with the U-value of the respective component obtaining an integrated indicator – the heat transmission power through the respective element; the causality of the latter with the total daily heating demand is very strong (Fig. 3c).

The need to compile the most comprehensive set of independent variables sometimes may lead to contradiction. To achieve the most objective computing KPIs from the selected set of indicators, as many quantitative indicators, which characterize very various aspects of the Process, as are accessible should be selected. On the other hand, to obtain reliable modelling results based on the causal relationships between the x independent variables and the dependent variable (Y) and to exclude individual deviations, a number of observation points k>>x is required. The stronger this inequality, the more accurate causal relationship (from the point of view of general causality) can be created. If k=<x, we could certainly find several relationships that perfectly reflect all the points of observation, but without the possibility of further generalization (which is needed for prediction and forecasting).

In cases with an insufficient number of observation points, an individual innovative approach is needed to overcome this shortage. So, using the fact that the DESI methodology had not been changed for several years, the data on EU countries for four years (2014-2017) were combined, thus virtually quadrupling the number of observation points (Karnitis et al., 2018).

In any case, contradictory conditions should be overcome.

3.2. Modelling

The choice of the optimal modelling method can be determined by the requirements set by the task to be solved: (1) the mode and tool of modelling should allow easy repeatability of the modelling if data and/or indicators change, and (2) the model should be implementable, recomputable and adaptable to a specific task by a person with mediocre programming skills.

Several general and specialized programming languages are suitable for our task. Among them is the well-developed and user-friendly **R** statistics environment, which is specialized for data analysis; it has a number of advantages that are important to our aim: (1) open access to the most popular operating systems; (2) a qualitative connected open integrated development environment (IDE) *RStudio*; (3) an interpretation mode that speeds up program development; (4) many open access external libraries for data analysis and display, and (5) simple and easy-to-learn syntax.

R allows for interactive execution of commands on the fly and an immediate display of the result; this speeds up the ad hoc analysis of data. It is possible to create complex data analysis programs. *RStudio* is a user-friendly environment for development that allows users to interact with, as well as develop and operate complex applications. For these reasons, we used the R statistics language, and used the built-in linear regression model, which is called by command *lm*.

3.3. Linear modelling: KPI computing

The multiple linear regression algorithm was chosen as the first option, because (1) it is mathematically the simplest method, (2) the obtained model is a linear expression, there is a simple and clear interpretation of the model, (3) basic knowledge in mathematics and programming is sufficient for model computation, and (4) using the mathematical expression of the model it can be easily calculated by spreadsheet or even by calculator. The linear regression presents the model in the form of a simple linear equation that shows well the effect of each KPI on the dependent variable, which characterizes the overall process.

R language contains the built-in linear regression model, which is called by command lm. The achieved level of compliance of the modelled target values (Ym) to the real values (Y) (mutual correlation) is serving as a quality criterion in the modelling procedure. The determination coefficient R2 is used as the unit of correlation during the modelling. Environment R issues a modelling result showing, inter alia, the estimated weighting coefficients βx and p-value for each indicator (see Table 1).

The first modelling procedure is performed including all indicators ix. Using a linear algorithm, the general mathematical function (f) is expanded in linear expression:

$$Ym1 = \propto + \beta1 \times i1 + \beta2 \times i2 + \cdots + \beta x \times ix \tag{2}$$

where α is the intercept and βx is the modelled weight of the indicator ix in the linear model.

The post-modelling selection of statistically significant indicators (KPIs) can be done either manually or by supplementing the standard modelling procedure; it is based on two features:

- The KPI is by definition the driver of the function Ym1; so, the indicators ix should be selected, for which the coefficient βx has a sign that drives the progress of Ym1: a (+) sign if the increase of ix promotes an increase of Ym1, but a (–) sign if the increase of Ym1 is promoted by a decrease of ix.
- Indicators ix should be selected, for which the p-value is less than a certain cut-off value; usually 0.05 or 0.1 is used as a threshold. It should be noted that other values also may be used.

Modelling procedure is repeated with only those indicators that have both features; the others relate to factors that are insignificant and even burdensome for the progress of Process. It is possible that in this narrower set of indicators some of them have lost statistical significance. They can be discarded and the next modelling procedure performed. After repeating the modelling procedure several times, we obtain an expression of resulting Ymlin, in which p-values of all indicators are small, while correlation between the (Ymlin) and (Y) remains strong (e.g., outputs of three-stage linear modelling of DESI 2019-2020 are shown in the Table 1). High statistical significance of these indicators shows their decisive role in the model's regularity; it clearly means that they are the sought-for KPIs (p1, p2,….pn). Action lines, which lead to progress in these KPIs, can be recommended to provide the overall progress of Process. Impact of other indicators on target variable (resp., progress of Process) is insignificant, even random; these indicators only complicate the model unnecessarily.

It should be noted that sometimes the general algorithms and standard software package should be adapted and supplemented to perform specific tasks. Thus, for modelling SD (Karnitis et al., 2021), a combined multi-stage modelling process was developed.

All 49 gathered indicators, which associate with one of the 13 sustainable development goals (SDGs) related to city-level performance, were grouped in 13 separate sets. All indicators were used for modelling as independent variables. According to the UN's vision of balanced SD, all SDGs should be evenly represented in the search for KPIs in the modelling process. Together, this means that only one indicator from the each set of indicators can be used as the predictor in each modelling procedure ($P = 13$ in SD study, providing $T >> P$).

Obviously, the highest obtained efficiency of the SD process will be achieved if all terms in expression (2) are compatible, and the progress of all predictors used had a positive effect on the progress of the target variable. To determine such good combinations of predictors, it was necessary to model all possible combinations, in which a single indicator from each indicator set was used as the predictor; more than 1.7 million models for subsequent testing were created in total. Each modelling procedure was supplemented with a compatibility test to check the mutual compatibility of the specific indicators' combinations and 232 "good" combinations were identified; they were analysed to determine statistically significant indicators.

Table1. Output data in DESI 2019-2020 multistage linear modelling.
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1st modelling procedure			2nd modelling procedure			3rd modelling procedure		
Indic.	Estimate	p-value	Indic.	Estimate	p-value	Indic.	Estimate	p-value
1a1	2.254e-03	0.350947***	1b1	0.0074910	0.02474 *	1b1	0.0094988	0.00276 **
1a2	-3.105e-04	0.732016	2b1	0.0006251	0.03654 *	2b1	0.0007550	0.00530 **
1a3	-7.071e-04	0.758560	3b2	0.0006766	0.10212	3b3	0.0016639	0.00647 **
1b1	1.809e-02	0.008904 **	3b3	0.0012171	0.05697 .	4a1	0.0005243	0.00772 **
1b2	7.449e-04	0.371116	3b7	0.0001773	0.34889	4a5	0.0968844	2.19e-05 ***
1b3	-1.958e-04	0.666097	4a1	0.0004773	0.01450 *			
1b4	-6.329e-03	0.008443 **	4a5	0.0770913	0.00122 **			
2a1	-3.410e-04	0.457984						
2a2	3.999e-04	0.147012						
2a3	8.261e-04	0.440479						
2b1	8.462e-04	0.010701 *						
2b2	-2.153e-04	0.048438 *						
2c1	-4.064e-03	0.090427 .						
2c2	6.483e-05	0.631385						
2c3	NA	NA						
2c4	-2.749e-04	0.586070						
2d1	-6.899e-05	0.848519						
3a1	-5.490e-05	0.941947						
3b1	-7.426e-04	0.071429 .						
3b2	1.311e-03	0.001997 **						
3b3	2.456e-03	0.000141 ***						
3b4	-3.193e-03	0.000178 ***						
3b5	-4.795e-04	0.216064						
3b6	-3.079e-04	0.411530						
3b7	7.587e-04	0.004916 **						
3c1	1.574e-03	0.118141						
3c2	5.645e-04	0.642668						
3c3	2.141e-04	0.854528						
4a1	4.479e-04	0.054915 .						
4a2	-1.628e-04	0.487793						
4a3	4.841e-04	0.142318						
4a4	-5.864e-04	0.105467						
4a5	8.120e-02	0.001513 **						

It should be noted that sometimes the general algorithms and standard software package should be adapted and supplemented to perform specific tasks. Thus, for modelling SD (Karnitis et al., 2021), a combined multi-stage modelling process was developed.

All 49 gathered indicators, which associate with one of the 13 sustainable development goals (SDGs) related to city-level performance, were grouped in 13 separate sets. All indicators were used for modelling as independent variables. According to the UN's vision of balanced SD, all SDGs should be evenly represented in the search for KPIs in the modelling process. Together, this means that only one indicator from the each set of indicators can be used as the predictor in each modelling procedure ($P = 13$ in SD study, providing $T >> P$).

Obviously, the highest obtained efficiency of the SD process will be achieved if all terms in expression (2) are compatible, and the progress of all predictors used had a positive effect on the progress of the target variable. To determine such good combinations of predictors, it was necessary to model all possible combinations, in which a single indicator from each indicator set was used as the predictor; more than 1.7 million models for subsequent testing were created in total. Each modelling procedure was supplemented with a compatibility test to check the mutual compatibility of the specific indicators' combinations and 232 "good" combinations were identified; they were analysed to determine statistically significant indicators.

Of the 49 indicators, 15 were found to be statistically significant in at least in one of the good combinations, even using a low threshold for evaluating the significance of predictors: the probability of rejection of the null hypothesis (p-value) is less than 0.1. These indicators divided into two drastically different groups by the intensity of their inclusions in the good combinations: 11 indicators were represented fewer than 10 times each, while 6 indicators more than 25 times each. Using only these 6 indicators as predictors we performed and individual modelling procedure, creating the mathematical expression of the high-quality 6-predictors' linear model (see Section 4).

3.4. Adding nonlinearity: quantitative clarifications

High numerical characteristics have been achieved creating linear models. Neverthe-less sometimes a detailed post-modelling analysis of residuals points toward an in-complete compliance of the actual data with the linear model. R diagnostic plots (frequency of residuals, residuals vs fitted, Q-Q, etc.) can show that the linear model does not fully capture the existing nonlinear relationship between the target variable and KPIs. E.g.,

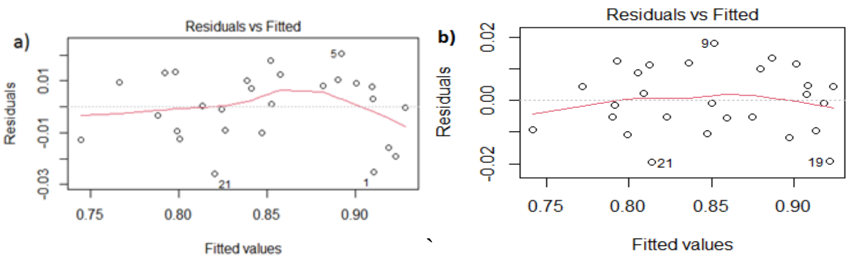

Figure 5. Residual characteristic of the urban SD (a) linear model, (b) final non-linear model.

the R diagnostic plot of the linear urban SD model (Karnitis et al., 2021) (Fig. 5a) shows that residuals are not relatively evenly spread around a horizontal zero line (especially at high fitted values).

It is possible to improve the model by adding nonlinearity to obtain the stronger causality and to specify the impact of KPIs on the target variable, i.e., KPIs weighting. This can be done both at the level of KPIs and the level of the mathematical expression of the model. We used the *RStudio* NLS function, which determines the nonlinear (weighted) least-squares estimates of the parameters of the nonlinear model.

The individual causal relationships between each KPI (pn) and the target variable (Y) were checked to process the level of KPIs. The real impact of specific KPI on (Y) is, of course, different from the individual regularity (e.g., due to some mutual impact of KPIs). Nevertheless, the strong qualitative difference of some individual (e.g., p2) causal relationship from the optimum linear one indicates a reduced quality of the linear model.

E.g., the strongly nonlinear data set p2 was tested by several relative nonlinear relationships, finding the fN(p2), which provides the strongest possible correlation between the modified Np2= fN(p2) and Y. By repeating the linear modelling, a corrected version of the expression (2) was obtained:

$$Ym1= \propto+\beta1\times p1+\beta2\times fN(p2)+\ldots+\beta n\times pn \tag{3}$$

In this way in the case of urban SD the G6.2 data set (Fig. 5a) was transformed into a new data set GN6.2, achieving a stronger linear relationship between GN6.2 and AHDI. The strongest correlation (Fig. 5b) was achieved by modelling the inverse proportionality expression.

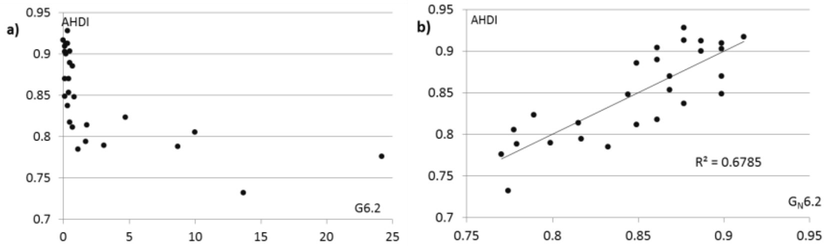

Figure 6. (a) Regularity G6.2. vs AHDI and (b) regularity G_N6.2 vs AHDI.

To process the level of the mathematical expression of the model the causal relationship between the Ymlin and the target variable (Y) was checked. It needs to be clarified whether the linear trendline provides the highest possible correlation, or whether there is some nonlinear function fP that improves the correlation. If so, the quantitative parameters of the model can be refined by modifying expression (3) accordingly, obtaining the final nonlinear expression of the model (Ym):

$$Ym= fP(\propto+\beta1\times p1+\beta2\times fN(p2)+\cdots+\beta n\times pn) \tag{4}$$

Thus, a corresponding scatterplot in (Karnitis G. et al., 2018) shows that the linear trendline of the data points is slightly skewed with respect to the axis of symmetry in case of urban SD. As a result, smaller fitted values of AHDIlin are generally slightly above the corresponding AHDI values, while large fitted values are below them. Such shifts indicate that the sigmoidal function is best-suited for improving the model. Several S-shaped functions were checked to decrease the aforementioned offset.

The R plot of the final nonlinear model (Fig. 5b) shows an improvement of the model quality in comparison with the one depicted in fig. 5a due to adding nonlinearity. Thereby, a more exact weighting of KPIs has been achieved.

3.5. Stability and sustainability testing

As each Process evolves, the numerical values of KPIs are changing in observation points. In addition to these justified changes, data holders are typically revising and updating the raw data repeatedly in order to reflect reality more accurately. We checked if and how these data changes affect the Process model, with respect to whether the model is stable against such data variability.

The model of EU economies was created for 2008-2015, that was a period of hard economic crisis, post-crisis recovery and return to sustainable growth (Karnitis et al., 2017); there were very different preconditions for the progress in particular years of the period. Calculations, which were made for each year using the created joint regularity, show a gradual increasing of correlation (Fig. 7a); it reflects the sustainability of both EU innovation-driven economic policy and the model, as well as an increasing convergence of the EU economics on the innovation-driven growth strategy. The fact that a single functional regularity could be used for years of such diversity, even that the regularity of deep crisis year 2008 could be used for growth year 2014 (Fig. 7b), attests to the universality and sustainability of the model.

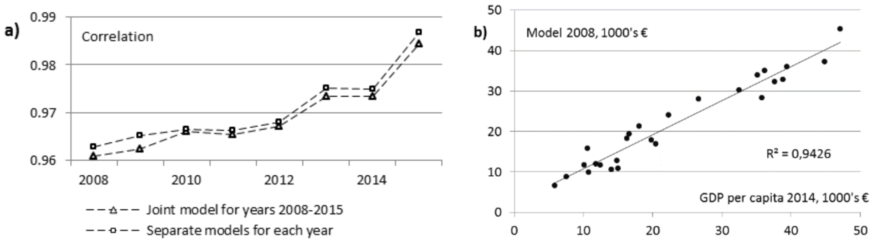

Figure 7. (a) Correlations of fitted GDP pc models 2008-2015 and (b) coincidence of GDP pc 2014 vs model 2008.

To quantify the stability of the urban SD model (Karnitis et al., 2021) several AHDIm5% control models were computed using the modified data, with 5% chosen as the maximum level of random variability of the input data; 5% is close to the median change in EU27 data over the previous 3 years. Of course, these changes have a corresponding effect on the modelled assessment of each individual country; however, the shift using models AHDIm and AHDIm5% does not exceed 0.2% across the EU27 countries.

Both cases clearly show that for practical applications annual calibration of models is not necessary.

In the DESI case the direct quantitative comparison of models 2014-2017 and 2019-2020 is impossible; nevertheless a qualitative comparison well reflects the compatibility of both models. European Commission in 2021 strongly adjusted the approach to DESI to reflect technological developments and the major political initiatives that impact sustainable digital transformation of EU countries, namely, Green Deal, Recovery and Resilience Facility and the 2030 Digital Decade Compass. The target function is extended from economic to socioeconomic development. An increased number of indicators is concentrated around four principal and interconnected policy areas: human capital, connectivity, integration of digital technology and digital public services. Particular attention is given to the fact that developments in DESI assessment should be achieved through improvements in all four areas. It was worth to repeat DESI modelling according to the updated indicator set and to compare the KPI selection with 2014-2017 for aforementioned policy areas; the data of 2019-2020 was used to avoid a possible Covid-19 impact in 2021.

Inclusion of the indicator “ICT specialists” in both KPI scopes clearly indicates that a lot of skilled ICT workers are necessary for socioeconomic development. As in 2017, the deficiency of ICT workforce remains a pan-European long-term problem. On the other hand, Internet use in 2020 in the EU27 countries has already become a common practice in society and therefore it has lost its statistical significance.

The connectivity to broadband networks (BB) should be evaluated as very valuable enabler of digital transformation. Both our models recognized fixed BB as a key technology, but technological development is perfectly reflected. Overall fixed BB coverage was important in 2014-2017, while due to technological developments exactly fast BB (NGA and VHCN) coverage became essential in 2021. On the other hand, the average level of the fast BB coverage in EU27 has achieved 85% of the overall fixed BB coverage; therefore the overall fixed BB coverage is no longer statistically significant indicator and it was excluded from DESI structure [40]. Despite the good availability of spectrum and stable y-o-y growth of the mobile data traffic, mobile BB still has a small weight in total broadband. 5G is still a technology of the future that has no significant impact on socioeconomic development today.

Both the big data (large data massifs) and open data (general availability of data) reflect the principles that have become strong enablers of any application and service in both private and public sectors. They are a basis for business transactions; they determine the current rapid spread of e-governance. Their selection justifies the exclusion of the narrower parameter “Electronic information sharing” from the set of indicators at the current level of development.

Sharply growing e-governance services on all levels (transnational, national, municipal) is reflected by radically increased statistical significance of e-government usage in 2019-2020.

In general, despite the drastic changes in the set of indicators, the heritability of the choice of KPIs is visible; they represent all four policy areas in DESI 2021 approach.

4. Results and conclusions

The achieved accuracy of the Processes’ linear models, excellent correlations with the real Processes and microscopic p-values of the models (Table 2) clearly demonstrate the correctness of the trend of the current study and practical perspective of obtained results and conclusions. Statistical significance of mathematically selected KPIs confirms that these KPIs are the real drivers of the Processes.

Of course, an always linear causal relationship between the KPI and the process variable cannot be expected; the location of the residuals in the R diagnostic plots is a reliable indicator that model can be further refined, reducing the effect of existing nonlinear causalities. One can see that by adding nonlinearity in the model numerical characteristics are further improved (Table 3); the nonlinear model coincides better with the actual values of target indicator in comparison with the linear one. Thus, a more exact weighting of KPIs has been achieved.

Table 2. Parameters of the Processes’ models.

Model type	Parameter	GDP pc 2008-2015	DESI 2014-2017	Heat transmission costs 2017	CSDI 2019	DESI 2019-2020
Linear	Indicators (x)	25	31	5	49	33
	KPIs (n)	3	7	2	6	5
	Coefficient of determination R^2	0.9305	0.766	0.9636	0.9421	0.7506
	p-value	<2.2E-16	<2.2E-16	5.9E-13	2.5E-11	2.1E-13
Non-lin.	Coefficient of determination R^2			0.9747	0.9638	
	p-value			7.9E-16	<2.2E-16	

Varied tests show stability and sustainability of the models. For practical applications annual update of the KPI set due to technological and/or economic development is not necessary; this needs to be done with significant changes of the set of indicators ix (e.g., due to differences in DESI 2015 and DESI 2020 methodologies).

Table 3. Numerical characteristics of linear and non-linear urban SD models.

Model	R^2	p-value	Residual standard error	Number of residuals, whose value		
				>2.5%	1.5-2.5%	<1.5%
Linear	0.9421	2.5×10^{-11}	0.01472	2	8	17
Non-linear	0.9638	$<2.2\times10^{-16}$	0.01042	0	3	24

The developed universal methodology is a reasonable compromise between accuracy, stability, and simplicity, which is a strong advantage for the practical applica-tions. KPIs and their weighting, found in this way, are mathematically computed, and modelling quality (accuracy) is quantifiable. Likewise modelling takes into account the complicated crosslinks between KPIs, as well as the integrity and interplay of separate action lines. Despite of the small number of selected KPIs that are the key drivers of the corresponding Process, the correlations between the modelled and the actual values of the target indicator are very, even extremely strong. The convincing results have obtained applying the mathematical calculation of KPIs; it shows a huge advantage over the methodologies using the voluntarily selected set of indicators and their weighting.

The popular Pareto principle, which over time has even been called "universal truth", states that typically 20% of inputs determine 80% of outputs. The computed pilot projects show that the described universal KPI computation methodology provides an even better outcome. While the average number of KPIs is around 20% of number of independent indicators (Table 2), the average degree of variability in the target variables that is explained by the KPIs (i.e., R2) is close to 0.9.

The methodology is applicable for variety of Processes; currently we are working on creation of models of the urban heating system for its management and renovation programs to reduce CO2 emissions and to achieve the climate goals, set by EU package "Fit for 55".

Only mediocre programming skills are needed for the application of the methodology. Its transparency, detailed description and open access to mathematical expressions computed for reported cases provide an opportunity for Process management to put the methodology into practice in order to monitor the achieved development level of the Process in statics and dynamics, to see progress and backwardness in particular aspects. It should be mentioned that KPIs reflect relevant action lines. This can be taken into account in strategic planning and decision-making; it will undoubtedly constitute a significant contribution to the Process development and management.

Based on the results of the study, the answer to the introductory question can be formulated, as follows: yes, it is always desirable to obtain more data that provide new data series for a comprehensive description of a particular Process from different aspects and for increasing the number of observation points. For the practical use of large amounts of data, it is recommended to find the KPIs driving the Process, using a mathematical algorithm that provides an objective choice of KPIs. The larger the amount of input data, the more accurate is the computation of KPIs and their weighting.

Based on the results of the study, the answer to the introductory question can be formulated: yes, it is always desirable to obtain more data that provide new data series for a comprehensive description of a particular Process from different aspects and for increasing number of observation points. For the practical use of large amounts of data, it is recommended to find the KPIs driving the Process using a mathematical algorithm that provides an objective choice of indicators. The larger the amount of input data, the more accurate it is to define KPIs and their weighting. At the same time it can be predicted that it will not be achievable the scope of KPIs that is fully adequate to the Process; selection of KPIs is a simplification that cannot includes all nuances of the Process. The model cannot be also an abstract representation of the data massif. The analyst always should carefully justify if and why a relationship between two variables has causality, do the existing relationships have predictive power for a new ones.

Acknowledgment

The study has been supported by the Latvian Council of Science project lzp-2021/1-0108 "Sustainable management of the urban heating system under EU Fit for 55 package: research and development of the methodology and tool".

References

Cakula, S., Krumins, K. (2020). Input determination for models used in predicting student performance. *Baltic J. Modern Computing*, **8**(1), 154–163. https://doi.org/10.22364/bjmc.2020.8. 1.08

Eggenberger, C., Rinawi, M., Backes-Gellner, U. (2018). Occupational specificity: A new measurement based on training curricula and its effect on labor market outcomes. *Labour Economics*, **51**, 97-107. https://doi.org/10.1016/j.labeco.2017.11.010

Gabbi, G., Matthias, M., Patrizi, N., Federico, P., Bastianoni, S. (2021). The biocapacity adjusted economic growth. Developing a new indicator. *Ecological Indicators*, **122**, 107318. https://doi.org/10.1016/j.ecolind.2020.107318

Goldstein, E.J. (2022). More data, more problems? Incompatible uncertainty in Indonesia's climate change mitigation projects. *Geoforum* 132. https://doi.org/10.1016/j.geoforum.2021. 11.007

Jiang, S., Shi, H., Lin, W,, Liu, H. (2020). A large group linguistic Z-DEMATEL approach for identifying key performance indicators in hospital performance management. *Appl. Soft Comput. J.*, **86**, 105900 https://doi.org/10.1016/j.asoc.2019.105900

You, J.W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet Higher Educ.*, **29**, 23–30. https://doi.org/10.1016/J.IHEDUC. 2015.11.003

Karnitis, E., Bicevskis, J., Karnitis, G., Pukis, M. (2021). Sustainable development model of EU cities compliant with UN settings. *Mathematics*, **9**(22), 2888 https://doi.org/10.3390/math9222888

Karnitis, G., Bicevskis, J., Virtmanis, A., Karnitis, E. (2022). Universal Methodology for Objective Determination of Key Performance Indicators of Socioeconomic Processes. In Proceedings of Digital Business and Intelligent Systems 15th International Baltic Conference, Baltic DB&IS 2022, Riga, Latvia, pp. 47-62.

Karnitis, G., Karnitis, E. (2017). Sustainable growth of EU economies and Baltic context: characteristics and modelling. *J. Int. Stud*., **10**(1), 209–224. https://doi.org/10.14254/2071-8330.2017/10-1/15

Karnitis, G., Virtmanis, A., Karnitis, E. (2018). Key drivers of digitalization; EU context and Baltic case. *Baltic J. Modern Computing*, **7**(1), 70–85. https://doi.org/10.22364/bjmc.2018.7.1.06

Kaur, M., Hewage, K., Sadiq, R. (2021). Integrated level of service index for buried water infrastructure: selection and development of performance indicators. *Sustain. Cities Soc*., **68**, 102799 https://doi.org/10.1016/j.scs.2021.102799

Kibira, D., Brundage, P., Feng, C., Morris, Kc. (2018). Procedure for selecting key performance indicators for sustainable manufacturing. *J. Manuf. Sci. Eng*., **140**(1), 011005 https://doi.org/10.1115/1.403 7439

Kubiszewski, I., Mulder, K., Jarvis, D., Costanza, R. (2021). Toward better measurement of sustainable development and wellbeing: a small number of SDG indicators reliably predict life satisfaction. *Sustain. Dev*., **30**, 139–146. https://doi.org/10.1002/sd2234

Li, Y., Gu, Y., Liu, C. (2018). Prioritising performance indicators for sustainable construction and development of university campuses using an integrated assessment approach. *J. Clean. Prod.*, **202**, 959–968 https://doi.org/10.1016/j.jclepro.2018.08.217

Onel, T., Ayla, S., Keskin, L., Parlayan, C., Yiğitbaşı, T., Kolbaşı, B., Yelke, T., Ustabaş, T. (2019). Leptin in sperm analysis can be a new indicator. *Acta Histochemica*, **121**, 43-49. https://doi.org/10.1016/j.acthis.2018.10.006

Osmond, P., Corkery, L. (2017). Developing key sustainability indicators for assessing green infrastructure performance. *Procedia Eng.*, **180**, 146–156. https://doi.org/10. 1016/j.proeng.2017.04.174

Polakoff, S. (2020). Why small business can no longer ignore big data. Forbes https://www.forbes.com/sites/forbestechcouncil/2020/05/20/why-small-business-can-no-longer-ignore-big-data/

Roldan-Garcia, M., García-Nieto, J,, Maté, A., Trujillo, J., Aldana-Montes, J. (2021). Ontology-driven approach for KPI meta-modelling, selection and reasoning. *Management*, **58**, 102018. https://doi.org/10.1016/j.ijinfomgt.2019.10.003

Sarma, U., Karnitis, G., Karnitis, E., Bazbauers, G. (2020). Toward solutions for energy efficiency: modeling of district heating costs. In: Tvaronaviciene, M., Slusarczyk, B. (eds.) *Energy Transformation towards Sustainability*, vol. **1**, pp. 219–237. Elsevier, Amsterdam. https://doi.org/10.1016/B978-0-12-817688-7. 00011-2

Sarma, U., Karnitis, G., Zuters, J., Karnitis, E. (2019). District heating networks: enhancement of the efficiency. *Insights Regional Dev.*, **1**(3), 200–213. https://doi.org/10.9770/ird.2019.1.3(2)

Tesic, M., Hermans, E., Lipovac, K., Pešić, D. (2018). Identifying the most significant indicators of the total road safety performance index. *Acct. Anal. Prevent.*, **113**, 263–278. https://doi.org/10.1016/j.aap.2018.02.003

Trerotola, S.O., Roy-Chaudhury, P., Saad, T.F. (2021). Drug-coated balloon angioplasty in failing arteriovenous fistulas: more data, less clarity. *Am. J. Kidney Dis.*, **78**(1), 13–15. https:// doi.org/10.1053/j.ajkd.2021.02.331

Warrier, S., Rutter, E.M., Flores, K.B. (2022). Multitask neural networks for predicting bladder pressure with time series data. *Biomedical Signal Processing and Control*, **72**, 103298. https://doi.org/10.1016/j.bspc.2021.103298

Wauchope, H., Amano, T., Geldmann, J., Johnston, A., Simmons, B., Sutherland, W., Jones, J. (2021). Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution*, **36**(3) 196-205. https://doi.org/10.1016/j.tree.2020.11.001

WEB (a). CDN Homepage, https://cdn.prohardver.hu/dl/cnt/2015-02/116080/desimodszertan.pdf, last accessed 2022/02/28.

WEB (b). EC Homepage, https://digital-strategy.ec.europa.eu/en/policies/desi, last accessed 2022/02/28.

WEB (c). EC Homepage. https://ec.europa.eu/info/research-and-innovation/statistics/performance-indicators/european-innovation-scoreboard_en#european-innovation-scoreboard-2021, last accessed 2022/02/28.

WEB (d). World Economic Forum Homepage. https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf, last accessed 2022/02/13.

WEB (e). WIP omepage. Global Innovation Index. https://www.wipo.int/global_innovation_index/en/

WEB (f). Cross-Sectoral Coordination Center Homepage. https://www.pkc.gov.lv/sites/default/files/inline-files/NAP2027__ENG_3.pdf, last accessed 2022/02/28.

WEB (g). EC. 2030 Digital Compass: the European way for the Digital Decade. COM(2021) 118 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118

WEB (h). Conference bites. Google Zeitgeist 9/27/11. http://www.conferencebites.com/2011/09/google-zeitgeist-92711/

WEB (i). Chalmers University of Technology Homepage. Thorstrom, M. Applying machine learning to key performance indicators. https://publications

Zacepins, A., Kviesis, A., Komasilovs, V., Bumanis, N. (2019). Model for economic comparison of different transportation means in the smart city. *Baltic J. Modern Computing*, **7**(3), 354–363. https://doi.org/10.22364/bjmc.2019. 7.3.03

Zhang, Y., Wang, J., Lu, J., Wu, J. (2020). Antibiotic resistance genes might serve as new indicators for wastewater contamination of coastal waters: Spatial distribution and source apportionment of antibiotic resistance genes in a coastal bay. *Ecological Indicators*, **114**, 106299, https://doi.org/10.1016/j.ecolind.2020.106299

Zuters, J., Valeinis, J., Karnitis, G., Karnitis, E. (2016). Modelling of adequate costs of utilities services. In: Dregvaite, G., Damasevicius, R. (eds.) *ICIST 2016. CCIS*, vol. **639**, pp. 3–17. Springer, Cham. https://doi.org/10.1007/978-3-319-46254-7_1

WEB (a). CDN Homepage, https://cdn.prohardver.hu/dl/cnt/2015-02/116080/desimodszertan.pdf, last accessed 2022/02/28.

Received December 20, 2023, revised February 13, 2023, accepted March 2, 2023