A Method and IR4I Index Indicating the Readiness of Business Processes for Data Science Solutions

Maxim Shcherbakov[1(✉)], Peter P. Groumpos[2], and Alla Kravets[1]

[1] Volgograd State Technical University, Lenin Avenue 28, 400005 Volgograd, Russia
maxim.shcherbakov@vstu.ru
[2] University of Patras, Greece University Campus, 26504 Rio Achaia, Greece
groumpos@ece.upatras.gr
http://www.vstu.ru, https://www.upatras.gr/

Abstract. The study shows our findings regarding the initialization and implementation of data science projects in existed business processes. The index readiness for intelligence or IR4I is proposed as an indicator of understanding about the readiness of your business processes for data science solutions. The index is based on the min-min convolution of various indicators: (i) business processes maturity indicators, (ii) indicators of the level of automatization and digitalisation of business processes, (iii) extract - transform - load (ETL) processes maturity indicators, (iv) data science infrastructure and technological stacks maturity. A new method of the IR4I index calculation is provided and its contains of six steps. Use case is based on real world task related to daily electric energy consumption forecasting for daily demand ordering. This example shows the application of proposed method and possibilities for improvement of business processes towards its intelligence and efficiency.

Keywords: Data science projects · Business process analysis · Energy management

1 Introduction

Nowadays, the design of systems based on artificial intelligence, machine learning and a large amount of data processing is the hot topic for academic and practical society. The new class of system based on data-driven approaches is widely discussed [2,15,19]. This tendency is connected with the law of technical systems development, in particular, with the shift of human intervention to a higher or superior level. It is also known as a supervisory control [21]. At the same time, the areas of implementation of technologies based on intelligent data processing are primarily focused on well-described business processes. Nevertheless, these well-studied processes involve a human as an executor or as a decision

M. Shcherbakov — The reported study was partially supported by RFBR research projects 16-37-60066 mol_a_dk, and project MD-6964.2016.9.

A. Kravets et al. (Eds.): CIT&DS 2017, CCIS 754, pp. 21–34, 2017.
DOI: 10.1007/978-3-319-65551-2_2

maker. The accumulation of data regarding various situations or creating training sets in terms of machine learning allows us to talk about the possibility of replacing a human with a machine not only for routine operations implementation but for making decisions as well. In a case of large amount of information machines to work more precisely and without experts biases [11]. In addition, data mining allows detecting hidden insights in data that are useful for business or to implement predictive analytics for making preventive or proactive decisions. The second option is based on the pattern 'detect-forecast-decide-act' of proactive computing approach [5]. In this case, it is assumed that some adequate probabilistic model of the processes exists. The deployment of components for predictive analytics into existing business intelligent solutions is associated with high expectations of significant performance increasing. It may be a trap for decision makers on the high level as they expect the increasing quality of the functioning of enterprise business systems and get more profit.

There is a research question: how to evaluate the readiness of current business processes or enterprise business system for its improvement from point of view of data science? Hence, it is reasonable to define a certain indicator that allows evaluating the degree of readiness of current business processes for development of data science solutions. Also, the indicator should help to estimate the current state of readiness and find the ways for processes improving.

The paper has a deal with new findings of an index of readiness of business processes for data science solutions. The index is named IR4I which is stands for Index of Readiness for intelligence. This IR4I is calculated based on four groups indicators: (i) business processes maturity indicators, (ii) indicators of the level of automatization and digitalisation of business processes, (iii) extract - transform - load (ETL) processes maturity indicators, (iv) data science infrastructure and technological stacks maturity.

2 An Index and a Method

2.1 General Idea

The method for evaluation the readiness of business processes for the implementation of data science solutions consists of six steps. As a result of methodology, the value of index IR4I is calculated. Based on the index value, conclusions are drawn regarding possible scenarios for business process improvement for successful data science solution implementation. The index of readiness to intelligent system deployment is calculated based on four groups indicators.

1. The first group includes indicators which assess the degree of business processes formalisation. In other words, these indicators estimate the maturity of business processes from the management point of view.
2. The second group contains on indicators for detecting the maturity level of automatization and digitalisation of business processes. They show the current state of enterprise business systems.

3. Indicators in the third group estimate the maturity of extract - transform - load (ETL) processes for certain enterprise business systems.
4. The last list of indicators evaluates the quality of existed data science infrastructure and technological stacks maturity.

As a result, the IR4I is calculated based on indicators and min-min convolution operation. Figure 1 shows the scheme of proposed method.

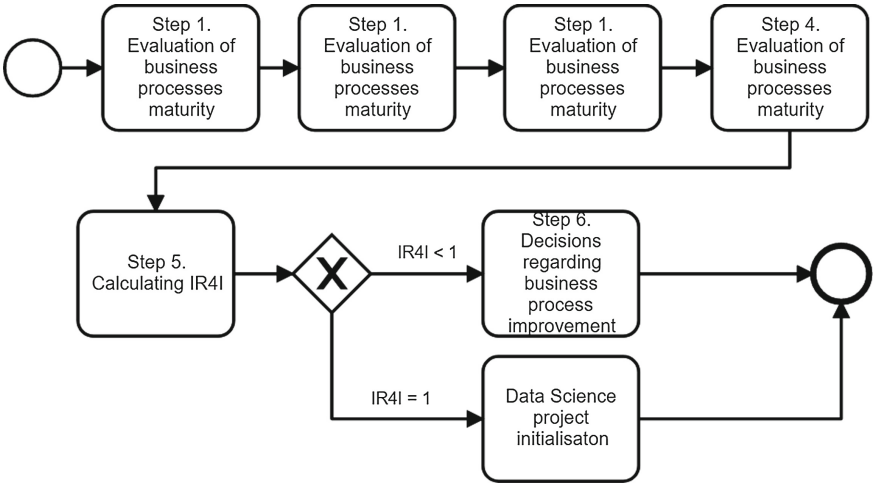

Fig. 1. A scheme for proposed method in BPMN notation. DS stands for data science, and TS is technology stack

Note, each indicator for a group has one out of three values: −1 means current state does not meet the requirements for data science solution development and implementation, 0 means current state satisfy requirements partially and 1 says about full satisfaction of requirements. This simplify way of calculation is considered as initial and can be improved using advanced techniques such as fuzzy cognitive maps [8,13].

2.2 Step 1. Evaluation of Business Processes Maturity

Before calculating of indicators, the business processes formalisation must be done. This is a standard step and is likely to be implemented in companies with an advanced level of maturity (not initial one where processes are unpredictable and poor defined). Business Process Model and Notation (BPMN) is reasonable to use for business processes formalisation and representation for further analysis. In the study, we focus on operations in business processes which are considered as checkpoint operations. It means there is a set of key performance indicators (KPI) defined and used for performance evaluation of the processes.

Checkpoint operations are subjects of evaluation of KPI's values. For such operations, KPIs are defined and formulas for KPI calculating are provided. Also, data sources containing necessary data for KPI calculating are specified. In this case, we can identify two benchmark values for business processes operations performance: (1) the desired KPI set by the manager at a higher level or driven by business requirements and (2) the mean or median KPI based on historical observations.

When the model of business processes was designed, this model is supplemented by two additional components. The first component is a structure of data flows indicating the input and output information for operations in a process's sequence. The second component is definition of data storages or data warehouses. For each data flow, a complete list of attributes for data obtained during the implementation of the operation is created. Each attribute is characterised by a name, a data type (numeric, categorical, text, images) and a range of values (optional). Data access attributes are defined for each data source:

- the type of access (manual, automatic, using API),
- the data access protocol,
- the request-response format,
- the implementation of data access mechanism e.g. data collectors, and
- access to data collectors.

The first group includes two indicators[1]. The first indicator $\alpha_1^{(1)}$ estimates the completeness of the description of business processes. The indicator has the following values according to the conditions

- $\alpha_1^{(1)} = -1$ if a business process is not described (using BMPN or another well-structured notation) or it is described partially, e.g. some operations are missed or key performance indicators are not specified,
- $\alpha_1^{(1)} = 0$ if data flows or data storages are not defined, e.g. there is no additional information about where data come from,
- $\alpha_1^{(1)} = 1$ in case of complete description of a business process.

The second indicator $\alpha_2^{(1)}$ shows the completeness of description of quality metrics or KPI for business process performance evaluation. It has the following values:

- $\alpha_2^{(1)} = -1$ if KPI is not defined or it is not assigned to a certain operation,
- $\alpha_2^{(1)} = 0$ if KPI does not have a formula for calculating or measurement tools are absent[2],
- $\alpha_2^{(1)} = 1$ if KPIs are defined completely.

Note, for $\alpha_2^{(1)}$ the typical KPI might be runtime or execution time. In this case it is reasonable to link execution time and business KPI related to the goals of companies.

[1] The superscripts shows the number of a group and subscripts is a number of an indicator in the group, e.g. $\alpha_j^{(i)}$ is j-th indicator in the i-th group.

[2] Also, there is no specific protocol for KPI evaluation.

2.3 Step 2. Evaluation of Automatisation Maturity

To determine maturity of business processes automatisation, it is necessary to classify the operations obtained model in the previous step. Classification can be performed based on types of operations in BPMN notation:

- tasks (actions) are operations having data as input and data as output in terms of data flow; input/output data might be represented as data frames;
- gateways (or decisions) are operations where choices made out a set of alternatives; the choice is made according to criteria;
- events (events) is an external influence on the internal process.

So the one feature inherent in all three types of activities can be single out. This feature is the existence of input and output data recorded and stored in a database. In terms of machine learning these records are considered as data sets for models training. Basically, the dataset contains on pairs of input and output vectors. From the standpoint of the readiness to implement data science solution, the existence of data sets is a crucial issue.

Consider the following indicators that are part of this group of indicators. The first indicator $\alpha_1^{(2)}$ estimates the degree of automation of business processes. The parameter takes the following values:

- $\alpha_1^{(2)} = -1$ if all actions related to information processing are performed manually;
- $\alpha_1^{(2)} = 0$ if at least one operation including information processing actions is performed manually;
- $\alpha_1^{(2)} = 1$ if all data processing operations in a business process is fully automated.

Note, in this research we consider the only activities having a deal with data or information processing.

The second indicator $\alpha_2^{(2)}$ evaluates the level of maturity for the activities related to quality evaluating procedures. The indicator might have one out three values:

- $\alpha_2^{(2)} = -1$ if values of more than one KPI are calculated or specified manually, or a manager participates in the data collecting process,
- $\alpha_2^{(2)} = 0$ if at least one KPI is calculated or set manually, or a human takes part in the collecting process and,
- $\alpha_2^{(2)} = 1$ if the quality indicators are determined automatically without a human intervention.

Finally, the third indicator $\alpha_3^{(2)}$ shows the availability of data in the format of datasets for further processing:

- $\alpha_3^{(2)} = -1$ if data does not represented as datasets (or dataframes),
- $\alpha_3^{(2)} = 0$ if data is partially represents as datasets (or dataframes) or some values are missed,
- $\alpha_3^{(2)} = 1$ if datasets (or dataframes) exist.

2.4 Step 3. Evaluation of ETL Process Maturity

ETL stands for a sequence of three types of operations over data: E – extract data from different data sources, T – transform data according to the storage structure requirements and L – load data to data storage or data warehouse. In this study, the aim is to evaluate readiness of all these operation for data science solution development and deployment. There are three indicators are considered here.

The first parameter $\alpha_1^{(3)}$ in a group characterises the maturity of the processes of data gathering from data sources. Note, the data source might be internal such as OLTP databases or external (e.g. weather forecast services). The indicator have the value

- $\alpha_1^{(3)} = -1$ if data collectors which collect data from various data sources are not exist,
- $\alpha_1^{(3)} = 0$ if data extracted manually,
- $\alpha_1^{(3)} = 1$ if data extraction is fully automated. Note, in the latter case, the schedule can be set for execution of load procedure [18].

The next indicator $\alpha_2^{(3)}$ is about the maturity of data transform process. The flowing values might be set

- $\alpha_2^{(3)} = -1$ if there are no any tools for data transform or data quality estimation,
- $\alpha_2^{(3)} = 0$ human intervention is required for data transform and quality estimation,
- $\alpha_2^{(3)} = 1$ the transform process is fully automated.

The third indicator $\alpha_3^{(3)}$ evaluates the maturity of data loading process into internal data storage for further analytics. Note, the data storage should keep raw data extracted from original data sources and transformed data according to the predefined scheme. The scheme is designed according to the requirements for further efficient data analysis. As an example, the solution can be built based on OLAP structure or HDFS with metadata storage. The indicator has one out of the three values:

- $\alpha_3^{(3)} = -1$ load tools are not available or not applicable,
- $\alpha_3^{(3)} = 0$ load requires intervention of a human,
- $\alpha_3^{(3)} = 1$ load is fully automated.

The main idea behind the indicator evaluation is a level of ETL process automation.

2.5 Step 4. Data Science Infrastructure and Technological Stacks Maturity Evaluation

Data science infrastructure is a complex of hardware and software available for generating statistical models or solutions based on machine learning approach. Note, this study uses CRISP-DM as a basic approach for data mining solution creating [3].

Consider the following indicators in this group. The indicator $\alpha_1^{(4)}$ assess the possibilities of reduction the performed in business processes task to commonly used statistical or machine learning problem. The interpretation of values for this indicator is following

- $\alpha_1^{(4)} = -1$ in case of untypical problem when it is difficult or unobvious to define a task statement,
- $\alpha_1^{(4)} = 0$ when task statement is possible to do, but it is unobvious how to reduce the task to a typical one. The term 'unobvious' means that data scientists should be involved in the processes,
- $\alpha_1^{(4)} = 1$ if the problem to solve reduce to the well-known task.

Note, the literature review shows various types of task classification [15]. The common thing of all different classifications is well-defined task statement.

One of the cost-intensive steps in the modelling process is the stage of data preprocessing or in other words preparation of data for modelling. The indicator $\alpha_2^{(4)}$ was introduced to estimate the preprocessing phase. The indicator takes the value

- $\alpha_2^{(4)} = -1$ if the preprocessing is not formalised and it is necessary to decide to make data preprocessing task statement and find methods to solve the task. Usually, $\alpha_2^{(4)} = -1$ if $\alpha_1^{(4)} = -1$,
- $\alpha_2^{(4)} = 0$ if human intervention in data preprocessing is required,
- $\alpha_2^{(4)} = 1$ if there is no need for preprocessing or preprocessing is performed automatically.

The next step is according to CRISP-DM is modelling stage [3]. It is necessary to define an indicator $\alpha_3^{(4)}$ characterising the degree of modelling automation. The indicator can be equal to one of the following values

- $\alpha_3^{(4)} = -1$ if the task is not defined, the model is not typical or modelling technologies are not defined,
- $\alpha_3^{(4)} = 0$ if an adaptation of new models (structural-parametric optimisation) is required,
- $\alpha_3^{(4)} = 1$ if the repeatable technology for modelling is developed.

The model performance evaluation is a mandatory stage of data science project. The indicator takes the value

- $\alpha_4^{(4)} = -1$ if there are no performance evaluation criteria,
- $\alpha_4^{(4)} = 0$ if it is required to adapt or test the quality criteria for applying,
- $\alpha_4^{(4)} = 1$ if the criteria are developed and experience is available.

2.6 Step 5. Calculating IR4I

The index is calculated according to the formula

$$IR4I = min_{i=1}^{n}\left(min_{j=1}^{k}\left(\alpha_j^{(i)}\right)\right), \tag{1}$$

where n is a number of operations in a business process, $\alpha_j^{(i)}$ – indicators mentioned in previous sections, k – a number of indicators for every operations. Assume, $\alpha_j^{(i)} \in \{-1, 0, 1\}, \forall i = \overline{1,n}, \forall j = \overline{1,k}$. The values of indicators can be interpreted as following:

- $IR4I = -1$ is a red zone does not meet requirements;
- $IR4I = 0$ is a yellow zone partially meet requirements;
- $IR4I = 1$ is a green zone, meet all requirements.

2.7 Step 6. Making Decision

The decision regarding the development and deploying data science solution for explored process is made. Also the choice of strategy for modification existed business processes is made in this step. Alternative strategies of actions are developed based on value of IR4I index and based on analyzing of included indications. So actions should increase the value of IR4I for certain business processes.

Analysis starts from the general overview of the IR4I calculated value

- $IR4I = -1$ indicates the low success rate for data science solutions implementation, so it is not reasonable to enrich existed automated system with data science components;
- $IR4I = 0$ tends to make additional analysis of the possibility (see horizontal and vertical strategies);
- $IR4I = 1$ is a green line, it is advisable to implement data science solutions.

3 Use Case

This section contains the results of application of proposed method based on new IR4I index. The use case describes the real-world problem of electricity consumption daily costs planning in an middle-size enterprise [12]. As a result of planning, the specific document is created which called *daily order* or *hourly bid*. The *daily order* containing a-day-ahead electric energy consumption demand in format of a set of raws (datetime, comsumption). Further, document sent to energy supply companies for further planning [12]. Based on the deviation of planned and actual electric energy consumption, the enterprise can be subject to a penalty in the amount according to the energy supply contract. The proposed in a paper method is applied for the certain process of daily electric energy consumption planning and daily order creating.

The middle-size manufacturing enterprise is considered as a object of energy consumption. Consequently, all indicators for IR4I are calculated for this specific case. Complementary to all calculated indicators values, the explanation is provided how calculating was made.

Figure 2 represents the business process of electric energy consumption planning in BMPN notation for middle-size manufacturing enterprise. For many countries, the special energy market exists. Specifically the marker deals with

the trade and supply of energy. If an enterprise is a participant of the energy market, the special everyday process of demand request generation is needed. The process starts according to the daily schedule.

The process starts with the Step 1 called 'Collecting of energy consumption data'. All information about electric energy consumption in previous time period is stored in plain table documents. The storage *sd1* is a folder with plain table documents. The output of the first step is a set of files containing data about energy consumption in previous time periods. If the folder *sd*4 contains files already, these files can be updated by new one. In the second Step 'Analysis of planned production plan', the information about planned production plan is gathered. The output of the step is files which represent union of planned production plan data with data about energy consumption in previous time periods. Finally, these files are stores on the folder *sd*4.

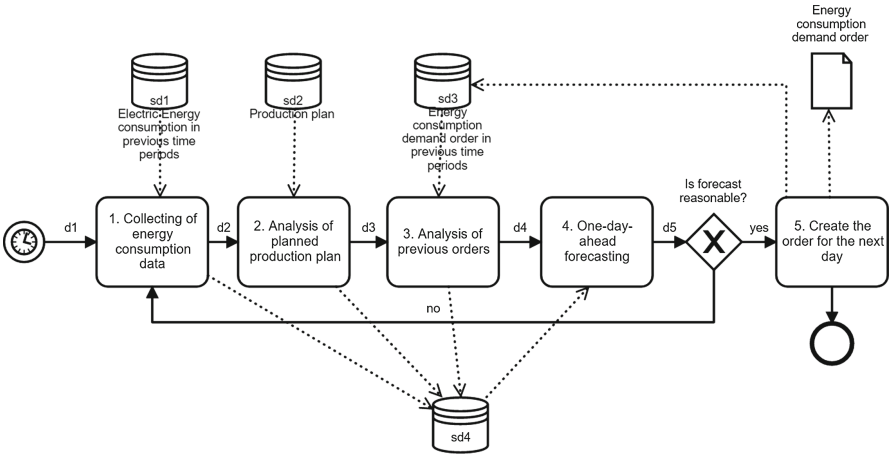

Fig. 2. A scheme for proposed method in BPMN notation

The third step titled as 'Analysis of previous orders' includes a procedure to collect data for further evaluation how data in orders varies from actual data, e.g. forecasting error evaluation. Based on data obtained in the previous steps, the forecasting of electric energy consumption is made in the step 'On-day-ahead forecasting'. The one-day-ahead forecast (KWh) is an output of the fourth step. As enterprise is a subject of fines in case of wrong demand ordering, the KPI for the processes following: mean absolute percentage error must not be exceed 15% and time for creating daily order must not exceed 2 h. The predicted values are estimated according to the manual verification procedure, and a manager who responsible for daily orders makes the decision about the finalisation of the process. The final step is creating the daily order document, signing and sending to the energy market broker.

Let consider the evaluation of described business process using declared method and IR4I index. The formalisation of the business process is high as

all necessary data is provided, so $\alpha_1^{(1)} = 1$. The completeness quality metrics or KPI description is high as well, as all necessary KPI was defined precisely (time and errors). The degree of automatisation of business processes $\alpha_1^{(2)} = 0$ as all procedures for data collecting are performed manually. The level of maturity for processes of data collecting and quality metrics evaluation is middle as well $\alpha_2^{(2)} = 0$; KPI related to forecasting performance is calculated manually. The indicator $\alpha_3^{(2)} = 0$ as data stored as plain tables with different format and it is necessary to make datasets for further processing.

Let analyse indicators of the second groups $\alpha_j^{(2)}$ and the actions for improvement of operations. Consider case when all data sources are defined and all data flows are represented in a formal way. In the domain we study it might be: *sd1* – is a data source of data about electric energy consumption. Formally, the data frames stored in *sd1* has the following structure

$$df_1 = \langle objectId, timeStamp, value \rangle ,$$

where *objectId* is a unique identification of the energy consumer, a *timeStamp* is a time where the measure was made, *value* is energy consumption value at the certain time stamp. Using the same formalisation, *sd2* contains data frames

$$df_2 = \langle resourceId, timeStamp, value \rangle ,$$

where *resourceId* – a unique identification of resource (energy consumer) using during the production. *sd3* includes data frames

$$df_3 = \langle objectId, timeStamp, orderedValue \rangle ,$$

where *orderedValue* is predicted energy consumption value for the future time stamps.

Data flows in Fig. 2 can be formalized as following

- $d1$ – is an event of initializations of the process;
- $d2$ – dataframe with the same attributes for df_1;
- $d3$ – contains two dataframes d1 and d2, where d2 has the same set of attributes as for df_2;
- $d4$ – contains three data frames, two from *d3* and a dataframe with the same set as for df_3;
- $d5$ – dataframe containing predicted values df_3 and a formal description of forecasting model md:

  $$md = \langle objectId, modelId, description \rangle ,$$

  where *modelId* is a unique id of a model and *description* is a high level description of the model, e.g. in *JSON* format.

In spite of data stored in electronic plain table format, $\alpha_1^{(2)} = 0$, as a manager intervention is needed for data collecting or extracting. A manager has to find a

necessary file in a folder *ds1* and open the file in appropriate software for further analysis. If there is no electronic storage or some information is stored physically on the paper (book of energy consumption), then $\alpha_1^{(2)} = -1$ (the worst case). If SCADA or an energy management automation system provide the automated formatted data downloading procedure, then $\alpha_1^{(2)} = 1$. In the use case, KPI's are not calculated automatically. Moreover, the operation numbered 3 is applied for this purpose. Based on the definition in previous section, the indicator $\alpha_2^{(2)} = 0$. Forecasting model performance should evaluate according to appropriate error measurements [1,17]. Note, if time measurement is performed manually it is not efficient due to the issues of human factor. Formally, data sets exist if all data is collected in the third steps in one place (data storage). We assume, that $\alpha_3^{(2)} = 1$ in this case. Basically, the data set should not be adapted for further processing. Also, the inner data is not enough for adequate modelling and an access to external data sets need to be configured. This is might be indicated by adding a new operation in data process.

Next, we evaluate indicators in the third group. For extract data indicator evaluation we set $\alpha_1^{(3)} = 0$ as all data source are defined, but data collected manually. In practice, these data are stored in copies of original data files gathered from data storaged *ds1*, *ds2* and *ds3*. Formats of initial data files obtaining from different data sources may varying. It is evident that, the specific format of data files is an additional limitation of the extract process. Next, we estimate $\alpha_2^{(3)} = 0$ because data transformation operation is required. The main reason for this is that gathered data is stored in files with a different format. A manager makes forecast using, for instance, electronic sheets but initial files stored in CSV format. Analogously, we set $\alpha_3^{(3)} = 0$ because of human intervention is needed in the data loading process. At least a manager should point the folder there the uploading data files are located. In fact, the manual operation required due to poor data quality, e.g. missing data or abnormal values. The data quality issues are crucial for loading processes as errors may occur. Finding the wrong data is the time-consuming operation.

Continuing, the estimation of existed infrastructure for data science solution implementation is made (indicators of the fourth group). As one-day-ahead electric energy consumption forecasting is the well-known problem, the problem is reduced to the time series forecasting task. Particular, this is a problem of n-ahead forecasting of univariate time series [10]. For our case study, the indicator $\alpha_1^{(4)} = 1$. The benchmark model for the task can be auto-regression model (AR, ARMA, ARIMA). Note, that time energy series forecasting can be done based on different approaches, e.g. probabilistic approach, symbol-based approach [6]. Data preprocessing is needed due to assumptions made in previous step. The typical preprocessing is sliding window approach. It means we set $\alpha_2^{(4)} = 0$. In the literature, we can find a lot of approaches for time series forecasting [7,12,14,20]. Special attention is given to existed automated forecasting approached and packages [9]. So, the indicator $\alpha_3^{(4)} = 1$. Forecasting performance evaluation can be made using various error measurements [17]. Traditionally, the

following error measurements are used for electric energy consumption forecasting: MAE, RMSE, MAPE [12,20]. However, due to error measurement drawbacks, the choice of appropriate criteria is subject to research [4]. Anyway, we estimate $\alpha_4^{(4)} = 1$.

As the result of estimating, the current use case has the following indicators represented in the Table 1.

Table 1. Indicators for use case

level	$i = 1$	$i = 2$	$i = 3$	$i = 4$
$\alpha_{(i)}^{(1)}$	0	1	-	-
$\alpha_{(i)}^{(2)}$	0	0	1	-
$\alpha_{(i)}^{(3)}$	0	0	0	-
$\alpha_{(i)}^{(4)}$	1	0	1	1

Based on the formula 1, the final value of IR4I is equal to zero. It means, we need to explore the process in detail and define a rational strategy for improvement (see Sect. 4). As the main conclusion, the development and deployment of data science solution for the defined business process is premature.

4 Discussion

So, if IR4I is equal to one, we conclude, that data science solution can be applied over existed automated systems. In this case the standard CRISP-DM process can be applyed for design this kind of solution. On contrast, if $IR4I < 1$ re-engendering of the business process need to be done.

Modification of the process can be performed according two strategies. Conditionally, the strategies are called "*horizontal*" and "*vertical*". The *horizontal* strategy is about the consistent improvement of operations at each level, from level 1 (formalisation of business processes) and ending the last level. Each modification is aimed at increasing the values of indicators. Until the indicator gets maximal value 1. The *vertical* strategy of changing involves the study of several processes in the business process.

Note that automation is a logical step in the development of data science solution. The components of data science solutions are considered as extensions to the existed automation system. Often (as in the example above), it does not make sense to create an additional analytic software architecture, but it is reasonable to use as extensions of existing one.

If a decision is made on the appropriateness of implementing data science solutions or components based on data analysis, the following indicators should be considered. Indicator of the model interpretability $\alpha_1^{(5)}$. The indicator takes the value -1 if the model is a black box and there are no whitening approaches

for the model. The parameter is set to 0 if the model is considered as a grey box or a black box with appropriate whitening methods. The proactivity of the processes $\alpha_2^{(5)}$, is characterised by the property of moving human functions to the supervisory level [16,21]. In this case, the indicator is equal to -1 if the model does not satisfy any of the properties and the human need to be involved in the implementation of the process. The indicator is set to 0 if at least one operation can be fully automated. Also, an indicator is equal to 1 if a human performs only supervision without permanent involving. The next indicator $\alpha_3^{(5)}$ - estimate the time performance for presentation of the result. Also, the indicator takes values -1 if the delivery time exceeds the time taken to perform the operation, 0 if the implementation of data science algorithms does not reduce the time of the person to perform the operation, and 1 if the time for performing is less before solution implementation.

5 Conclusion

The readiness of the current business processes for the implementation of data science solution can be formalised and evaluated. We propose the IR4I which is stands for Index of Readiness for intelligence is an index using the min-min convolution of various indicators including: (i) business processes maturity indicators, (ii) indicators of the level of automatization and digitalisation of business processes, (iii) extract - transform - load (ETL) processes maturity indicators, (iv) data science infrastructure and technological stacks maturity.

The method of evaluation is considered which is consist of six steps. The method and IR4I index can be used for the initial stage of data science solution design. However, the proposed IR4I is not sensitive for evolution of improvement. The future works should include the problem of interpretation of changes in business improvement process.

Acknowledgments. The reported study was partially supported by RFBR research projects 16-37-60066 mol a dk, and project MD-6964.2016.9. Also authors would like to thank Pavel Vorobkalov for fruitful discussion and anonymous reviewers for fruitful remarks.

References

1. Armstrong, J.S.: Evaluating forecasting methods. In: International Journal of Forecasting, vol. 30, pp. 443–472. Kluwer Academic Publishers, Norwell (2001)
2. Arnott, D., Pervan, G.: Eight key issues for the decision support systems discipline. Decis. Support Syst. **44**(3), 657–672 (2008)
3. CRISP-DM: Still the top methodology for analytics, data mining, or data science projects. http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html. Accessed 01 Apr 2017
4. Davydenko, A., Fildes, R.: Forecast error measures: critical review and practical recommendations. In: Business Forecasting: Practical Problems and Solutions. Wiley (2016)

5. Engel, Y., Etzion, O.: Towards proactive event-driven computing. In: Proceedings of the 5th ACM International Conference on Distributed Event-Based System, pp. 125–136 (2011)
6. Golubev, A., Shcherbakov, M., Shcherbakova, N.L., Kamaev, V.: Automatic multi-steps forecasting method for multi seasonal time series based on symbolic aggregate approximation and grid search approaches. J. Fundam. Appl. Sci. **8**(3S), 2529–2541 (2016)
7. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. Int. J. Forecast. **22**(3), 443–473 (2006)
8. Groumpos, P.P.: Fuzzy cognitive maps: basic theories and their application to complex systems. Fuzzy Cogn. Maps **247**, 1–22 (2010)
9. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. J. Stat. Softw. **27**(3), 1–22 (2008). doi:10.18637/jss.v027.i03. ISSN 1548-7660
10. Hyndman, R.J., Athanasopoulos, G.: Principles and Practice. OTexts, Melbourne (2013). http://otexts.org/fpp/
11. Kahneman, D.: Thinking, Fast and Slow. Farrar, Straus and Giroux, New York (2011)
12. Kamaev, V.A., Shcherbakov, M.V., Panchenko, D.P., Shcherbakova, N.L., Brebels, A.: Using connectionist systems for electric energy consumption forecasting in shopping centers. Autom. Remote Control **73**(6), 1075–1084 (2012)
13. Mamlook, R., Badran, O., Abdulhadi, E.: A fuzzy inference model for short-term load forecasting. Energ. Policy **37**(4), 1239–1248 (2009)
14. MIRACLE Consortium: Micro-Request-Based Aggregation. Forecasting and Scheduling of Energy Demand, Supply and Distribution (2010)
15. Nisbet, R., Elder, J., Miner, G. (eds.): Handbook of Statistical Analysis and Data Mining Applications. Academic Press, Cambridge (2009). ISBN 0123747651, 9780123747655
16. Salovaara, A., Oulasvirta, A.: A user-centric typology for proactive behaviors. In: Proceedings of the 3rd Nordic Conference on Human Computer Interaction NordiHCI, pp. 57–60. https://doi.org/10.1145/1028014.1028022
17. Shcherbakov, M.V., Brebels, A., Shcherbakova, N.L., Tyukov, A.P., Janovsky, T.A., Kamaev, V.A.: A survey of forecast error measures. World Appl. Sci. J. **24**(24), 171–176 (2013)
18. Sokolov, A., Tyukov, A., Sadovnikova, N., Zhuk, S., Khrzhanovskaya, O., Brebels, A.: Automatic information retrieval and preprocessing for energy management. In: Kravets, A., Shcherbakov, M., Kultsova, M., Shabalina, O. (eds.) Creativity in Intelligent, Technologies and Data Science: First Conference, CIT&DS 2015, Volgograd, Russia, September 15–17, 2015, Proceedings, pp. 462–473. Springer International Publishing, Cham (2015)
19. Stluka, P., Ma, K.: Data-driven decision support and its applications in the process industries. Comput. Aided Chem. Eng. **24**, 273–278 (2007)
20. Taylor, J.W., Espasa, A.: Energy forecasting. Int. J. Forecast. **24**(4), 561–565 (2008)
21. Tennenhouse, D.: Proactive computing. Commun. ACM **43**(5), 43–50 (2000)

http://www.springer.com/978-3-319-65550-5

Creativity in Intelligent Technologies and Data Science
Second Conference, CIT&DS 2017, Volgograd, Russia,
September 12-14, 2017, Proceedings
Kravets, A.; Shcherbakov, M.; Kultsova, M.; Groumpos,
P.P. (Eds.)
2017, XVII, 887 p. 335 illus., Softcover
ISBN: 978-3-319-65550-5