Applying data mining techniques to discover KPIs relationships in business process context

Emna Ammar El Hadj Amor[1]
[1]ISITCom Hammam Sousse, RIADI Laboratory-ENSI Manouba, Tunisia

Sonia Ayachi Ghannouchi[2]
[2]ISG Sousse, RIADI Laboratory-ENSI Manouba,

***Abstract*— Organizations need to continually improve and review their critical business processes. In addition, it is crucial not only to track the business process (BP) behavior and to derive key performance indicators (KPIs) but also to understand all necessary concepts and incorporate domain knowledge of the field. The purpose of this paper is to gain a deeper understanding of the interrelationships between all concepts and performance measurement raw data to extract their real meaning. In order to meet these challenges, first, we explore several qualitative and quantitative indicators for measuring the performance of BPs. Second, we develop a new ontology for the representation of these performance indicators. Then, we are based on data mining techniques to extract the most important information from data measurement and to discover all necessary relationships between indicators.**

Keywords- Business Process (BP), Process Improvement; HealthCare process; Performance measurement; Key Performance Indicators; Ontology; Data mining; association rules

I. INTRODUCTION

The improvement of business processes is essentially based on collecting all relevant indicators designed for both effective management and process improvement. Depending on the context, several kinds of Key Performance Indicators (KPI) can be developed, including quantitative or qualitative aspects, and different data sources can be used for retrieving and calculating their appropriate values. In this perspective, Kaskinen puts pressure on the point that it is hard to improve what cannot be measured [1]. Performance Indicators (KPIs) provide critical information to the organization for monitoring and predicting business performance in accordance with strategic objectives [2]. According to Parmenter [3], key performance indicators (KPI) represent a set of measures focusing on those aspects of organizational performance that are the most critical for the current and future success of the organization. Furthermore, business process management and continuous improvement initiatives can be significantly enhanced by real-time measurement and traceability information. Kronz points out that the basis for all process controlling is a process-oriented KPI system that links the process perspective to the essential controlling aspect of the business [5]. Usually, the data related to the business process execution is useful for monitoring facilities. To monitor KPIs, organizations rely on reports and dashboards [3] presenting one or more KPIs together in order to help decision-makers identify opportunities for improvements or re-engineering the desired process. However, this practice provides partial information to decision makers, without taking into account which activity in the BP is involved and which inter-relationships between KPIs are considered. There are many measurements that could be used in the evaluation of business processes. In general, we can deal with two different aspects: qualitative and quantitative. In each aspect, it's important to select the key measures that have a significant impact on the organization. In the same context, performance measurement has several practical issues, such as unavailability of performance measurement system, distribution of data in multiple locations, incompleteness, lack of traceability of the BP or they are expressed informally in natural language etc. usually the quantitative and qualitative KPIs data are recorded in separated locations (e.g. document, system logs, tables or spreadsheets) without establishing any link or relationship with other elements related to the same business process. This leads to several problems: difficulty to maintain the coherence across both types (quantitative and qualitative), e.g. what happens if an activity or a KPI is removed and there exists another KPI (in the same type or different type) that is defined only for that activity or depended from that indicators? What happens if a KPI value is changed and if it has a link which is defined by referencing to another indicator? Is this link still correct or does it still exist when changing the measurement instance data or the period of the analysis? What happens when an organization develops new KPIs or analyzes a new data source? Despite the difficulties in measuring performance indicators of business processes, another important difficulty is how to gather the existing information related to different concepts into useful practices. Therefore, there is currently a need to improve the KPI improvement process, providing decision-makers with information about the relationships among KPIs. To resolve this issue, Data mining has a great potential to enable healthcare systems to use data more efficiently and effectively [15]. The ability to use data in databases in order to extract useful information for quality health care is a key of success of healthcare institutions [16].

Furthermore, indicators are constantly under pressure to control rapidly the business process. As a result, it's crucial not only to analyze business processes for critical changes based on acceptable limits for key parameters (e.g. cycle time measurements exceeding a given target) but also to monitor changes in KPIs relationships data set to measure goal progress and identify areas needing adjustment. The research will include how to manage this gained KPIs data in order to efficiently discover hidden information. Data mining is conducted through the use of well-known data mining algorithms to extract association rules from the given data. The aim of our work is to study the efficiency of indicators and exploring improvement opportunities. In this view, we were motivated by a concrete case study of the healthcare process in the emergency department of "Farhat Hached" hospital in Sousse (Tunisia).

The remainder of the paper is organized as follows. Section 2 presents some related work in this area. Section 3 presents our KPI measurement approach. Section 4 discusses experimental results using data mining techniques. A conclusion ends this paper in the last Section.

II. Related Work

Many works applied the Apriori algorithm in the healthcare domain [16] [17] [18] [19]. For example, in [16] the author used this algorithm in order to find out the associations between diagnosis and treatments in medical billing data. In [18] this algorithm was applied to discover frequent diseases in medical data in particular geographical locations at a particular period of time.

Ortega [28], defined a catalog of analysis operations of PPIs that allow to automatically extracting implicit information from them and their relationships with the BP. In their work, they use Description Logics (DL) to implement analysis operations in order to infer the required knowledge from PPI definitions.

In [26] authors propose an ontology-based data mining approach which utilizes Apriori algorithm. At the end of the case study, they have determined the association rules for the smartphone users. In other words, they analyzed consumer preferences in terms of mobile OS usage. The knowledge can then be integrated into the decision support process (e.g. offer the right product/advertisement to the right consumer) of telecommunication companies' marketing departments.

In [27] the authors create a multidimensional model to support KPIs calculation and provide additional analytical capabilities. After that, they analyzed the candidate KPIs through data mining techniques to ensure that they reflect the relationships identified during the business strategy modeling. The Using of DM techniques to extract relevant KPIs proposed by authors consists of five processes: (1) Preprocessing, (2) Potential anomalies detection, (3) Difference series calculation, (4) Analysis of pair-wise relationships between series (e.g. correlation, time series analysis, and linear regression), and (5) Analysis of compound relationships (e.g. classification data mining technique).

In our approach, we explore all information that can be highly valuable in KPI selection, not only from business process logs but also from other documents (e.g. questionnaire, interviews) where the implications of the KPI data are unknown. Thus, eliciting their relationships and choosing the appropriate data set (e.g. data set according to the period of analysis) can make these data actionable and adding value to the business process improvement. In addition, we define key measurements required for analyzing business process (e.g.waiting times, execution times) and other related performance information. For example, to improve the overall performance of the healthcare business process, the evaluation focus on interesting quantitative KPIs related to the behavior of the process and also on interesting qualitative KPIs related to patients experience in the emergency department (ED). So, we need to identify relationships between KPIs to obtain relevant information that can assist BP analysts during BP improvement. This work suggests a novel approach based on association rules towards building interesting relationships between KPIs. Some benefits of the proposed approach are as follows: (1) Evaluating measurement taking into account both quantitative and qualitative aspect; (2) Involving all informed persons about the business process by concentrating on the most valuable indicators relationship based on their association rules; (3) Checking and analyzing KPIs behavior using existing data to improve the KPIs data or to improve the business process.

III. Bussiness process measurement

Observing the huge volumes of information about indicators for different purposes and domains and the heterogeneity of information sources, it becomes essential to provide mechanisms for structuring and retrieving useful and interesting data. The semantic representation of KPIs is a step towards KPI Information discovery and then KPI improvement. In fact, analyzing business processes requires computing measurement that can help determine the health of business activities and thus that of the whole department or organization.

To illustrate our approach we were motivated by the healthcare process because it is generally simultaneously characterized by information richness, but poor knowledge because of the lack of relevant data, unknown relationships between data and goals, conflicting KPIs, conflicting goals, and poorly understood measurement.

The business process in this case study shown in figure 1, begins by the registration activity. After that, in order to arrive at a preliminary conclusion about the status of the patient, a sorting activity represents the second task in the BP and the first point of contact with medical staff. This activity consists in recording the preliminary observations and prioritizing the patients according to their degree of urgency. The following tasks depend on the status of the patient. We find various cases of consultations such as consultation in the delayed emergency sectors in the case of non-urgent patients, consultation in the box (it can be a simple consultation or surgical consultation), consultation in the crash room (the patient is of serious harm which requires immediate medical attention) and finally the last case of consultation in the supervision room if the condition of the patient is not stable.

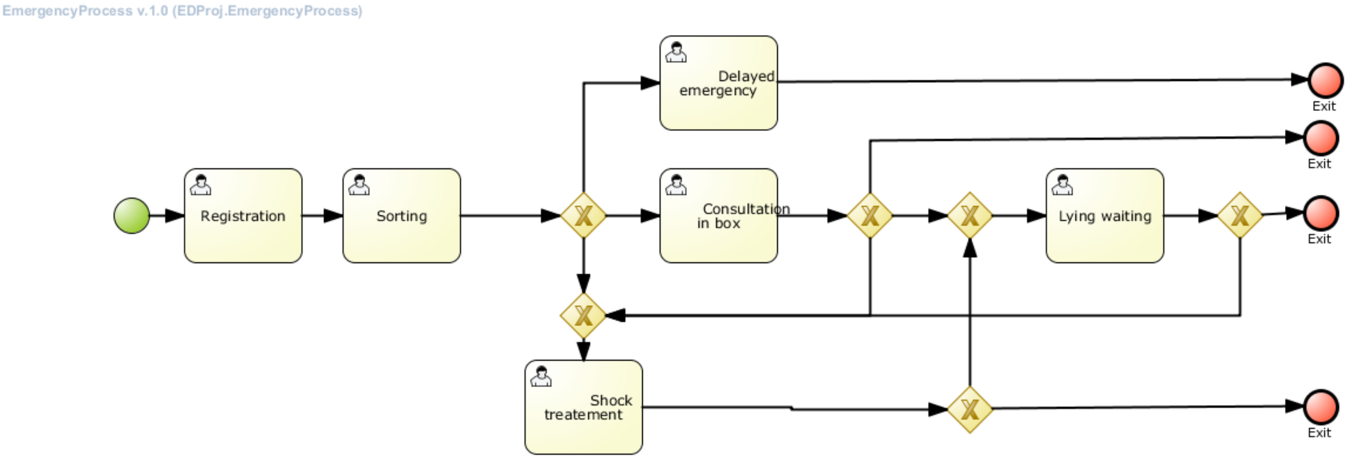


Figure 1. Health care business process

Typically, a business process is based on the workflow of activities which depend on the events that initiate the process, the participants in the process and process measurements. Event data can be analyzed according to different perspectives [23]: (1) the control-flow perspective; (2) the organizational perspective; (3) the data perspective; and (4) the performance perspective. The control- flow perspective is concerned with the process behavior, namely the activities in the process and their order of execution. The organizational perspective focuses on the relationships between the users who performed the activities, such as whether they belong to the same or to different groups or organizational units. The performance perspective aims at detecting bottlenecks or calculating performance indicators, such as throughput times and sojourn times. The data perspective is related to the data objects that serve as input and output for the activities in a case.

We refine Key Performance Indicators into two classes: basic KPIs (For example, the KPI Process Instance execution time is a basic metric which takes as input one Process

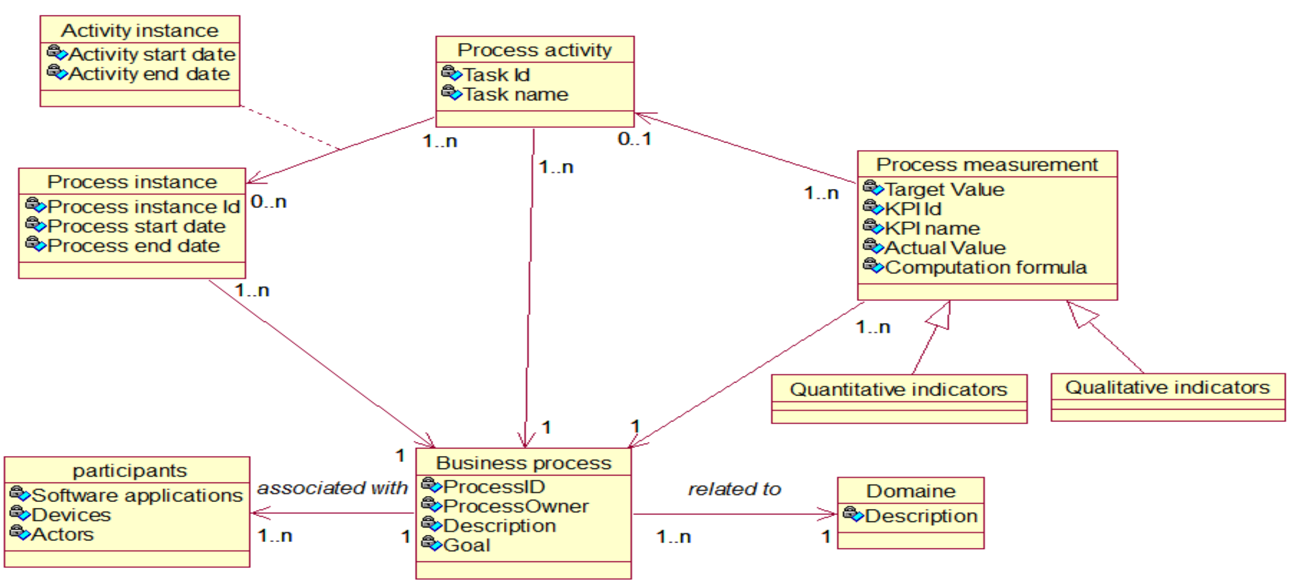


Figure 2. Proposed metamodel

Instance) and aggregated KPIs which are built on basic KPIs (For example, the min of total time spent in the ED by all patients (all process instances) in all activities). Additionally, in order to provide a convenient way for business analysts to analyze metrics, we take into consideration only the basic metrics because the number of individuals (e.g. patients, medical staff) of the same kind (e.g., a set of Process Instances) to be analyzed varies over time and then the overall perception of some aspects of interest changes (e.g. the satisfaction will be evaluated over each individual). Based on basic KPIs, we define quantitative indicators and qualitative indicators. We analyze both of qualitative and quantitative KPIs. In this work, in one hand, we extract relevant quantitative information from the event logs produced during the BPM execution and on the other hand, we extract relevant qualitative information related to healthcare domain by addressing patient queries. To build our ontology we were based on the proposed metamodel described in figure 2. We essentially focused on process tasks and qualitative and quantitative indicators, and we defined a set of individuals for specifying KPIs and activities because regarding the contextual characteristic of KPIs: what can be considered as a KPI within a particular domain might not be as relevant for another. In our ontology, we represent, on one hand, the relationships between an activity and the attached KPI, and on another hand, the relationships between indicators, we define a set of object properties. Therefore, we create three object properties related to our KPIs. For example, the property "related-with" might link the possible individuals from the qualitative class (KPI related to a patient satisfaction) to the related individuals in the quantitative class. Also, in order to represent the relationships between indicators in the same category, the owl model implements the properties "Depend-quantitative" and "Depend-qualitative" which indicate the links between quantitative/ qualitative indicators, their relationships and mainly represent the need to share data between them. Furthermore, properties can have inverses and they may be transitive or symmetric. For example, the inverse of "has-activity" is "has-indicator". Those definitions of relationships and the relative individuals need to be highlighted and modeled in order to aid the monitoring of KPIs that contribute to performance improvement. Moreover, we envision a collection of Datatype properties to describe relationships between individuals and data values.

According to literature, we can say that we have to know more about the real meaning and the implications of performance measurement value and to understand why an indicator is needed and what it is measuring and what decisions it supports. Thereafter, the business process of the domain is built and the relational data logs are analyzed. The database log instance classes in our proposed ontology contain two subclasses: the first one corresponds to the essential tables extracted from business process management system and the second one corresponds to real and estimated values related to our key performance indicators for all completed instances in the healthcare BP.

The jBPM_data concerns all necessary information collected during business process execution. The KPI_data concerns all instances of both quantitative and qualitative aspect. KPIs data mainly contains all the quantitative and qualitative measurement regarding patients. The storage of such type of data is variously depending on the period of analysis and the evaluation of organization performance. Due to the continuing increase of measurement healthcare data size, we focus on the analysis of the KPIs in the observation period.

This data will enable us to put together a comprehensive picture of the BP. The proposed business process models are deployed with the use of jBPM software, where 100 instances have been created through the execution of the process which can be statistically considered as significant and consequently it forms the basis for continuous process optimization. For this reason, we were essentially based on the availability of jBPMlog to extract all relevant data. jBPM is an open source BPMS. It features a robust management console and development tools, with user support during the business process lifecycle: development, deployment, and versioning [21]. Our choice of jBPM is motivated by its capability in answering healthcare performance measurement needs, such as the availability of history logs of all process instances as the data source of measurement. Persistent data is used for performance measurement. To persist jBPM historical data, in our work, we configured PostgreSQL database system to store the history log through its persistence module. Data involves huge volume and growing data sets. After the analysis of the available data in jBPM logs, we focus on three tables, described below, which are in charge of retrieving the data needed for our key performance indicators defined above:

- Processinstancelog table which contains the basic log information about a process instance. Every process instance has an identifier.
- The bamtasksummary table which contains summarized information about all process tasks.
- Variableinstancelog table which stores the data (the value) corresponding to the changes in process variables that are manipulated during process executions.

We also conclude that not all KPIs can be directly retrieved from these logs and that the representation of KPI values, especially in the calculation of waiting time and duration in milliseconds, can be misunderstandable for the decision maker. For this reason, we create another data table which contains all real values related to our research. The new data table named "real values KPI instance" contains all real values related to our research. All data are retrieved from KPI database using PostgreSQL. PostgreSQL is an open source object-relational database system [22]. KPI_data tables contain all information related to quantitative process indicators and qualitative indicators related to the same process instance id.

IV. KPI KNOWLEDGE DISCOVERY

As aforesaid, indicators can be vague and lack of semantics. Furthermore, selecting KPIs at the Emergency Department (ED) requires a high cooperation between its quantitative indicators and their related qualitative indicators. For this purpose, it is necessary to adequately link quantitative and qualitative measurement affecting task process, and design associated relations. The aim of this step is to guarantee a continuous improvement approach based on the analysis of completed instance.

The target in this phase is to find all interesting KPI that are linked together and approved by the majority of patients and hence should be linked in a proper way in order to maximize patient satisfaction and contribute to improving the BP. in this work we are based on 100 rows referring to patient data since an instance of BP represents a patient. Data in KPI table needs to be checked and evaluated by the organization to detect if KPI values are reaching the desired results. As a result, we create another table which contains all estimated values related to the patient instances (real value KPI instance classes). This table contains two values ("ok" codes the KPI value which is considered by the ED as acceptable and "not ok" corresponds to the fact that the KPI value is not acceptable). We found that learning ontology from existing information resources is a good solution to explicitly express the semantics of information sources. The main advantages of this table provide the basic information to the decision maker to execute corrective actions in case of important deviations.

Data Mining mainly extracts the meaningful KPIs data which were previously recorded in the Estimated KPIs table and derived from the event logs and the qualitative inquiry. This performance measurement can be then interpreted and translated into the knowledge where the discovering of interesting decisions can become possible.

In fact, to make any decision, it is important to analyze all the relevant row data of our performance measurement. For this purpose, we use data mining, especially association rule learning as a research method in this stage.

Agarwal and his colleagues at IBM Almaden Research Center introduced a novel association rule algorithm Apriori [24] where the association mining can be applied to real databases to extract association rules. The Apriori algorithm requires two user parameters configuration: the first one is support and the second one is confidence. These parameters are used to significantly limit the search for frequent itemsets. The choice of the thresholds minimum support and minimum confidence clearly influences the number and the length of rules generated by algorithms. In order to understand all KPIs interaction, we developed a framework to extract association rules from the KPI data. The main advantages of this framework are related on one hand to the evaluation of several performance measurements, and on the other hand to the advantages brought by the semantic technologies to improve the BP. In fact, to make any decision, it is important to analyze the knowledge model and all the relevant row data of the performance measurement by taking into account interrelationships and influences between them. The framework provides a comprehensive view of performance measurement and covers many types of measurements and the semantic link between them. Furthermore, it is intuitive, i.e. it is easy for users to understand and evaluate the performance of the considered BP. Hence, it will allow discovering relations between indicators. More precisely, first, it helps the decision maker to choose both quantitative and qualitative performance metrics in relation to a business process and then, the framework operates at a sufficiently detailed level such that it provides a comprehensive basis for assessing the relations between KPIs. Because of Java's robustness, we decided to code and implement the Apriori algorithm in our framework using NetBeans.

By using Apriori algorithm we aim to find frequent associations and correlations among sets of items. If a KPI value is not frequent, no association rules related to the KPIs are generated. Association rules mining algorithm needs to be configured before learning. So, we give in advance appropriate values for the parameters. Decreasing the minimum support and minimum confidence will increase the number of exploitable rules.

Figures 3 and 4 show some execution results obtained from our framework based on Apriori algorithm to predict the occurrence of an item based on the occurrences of other items in the transaction.

For instance, in figure 3, we choose two Indicators (Quali_KPI2 and Quali_KPI8) linked by "depend_qualitative" object property. In figure 4, we choose two Indicators (Quali_KPI11 and Quanti_KPI17) linked by "related with" object property.

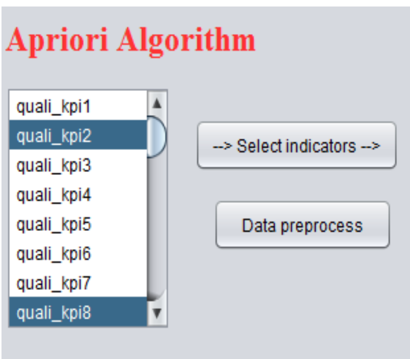
In order to validate our results, we use WEKA data mining tools with the Apriori algorithm where the same input data was applied. Figures 5 and 6 show the result of examples 1 and 2.

The confidence of the first rule in Weka results that contain both quail_KPI8 (The quality of care for patients by medical staff) and quail KPI2 (interest and attention brought by medical staff) is 0.89. A confidence of 89% means that 89% of the patients who were satisfied with quail_kpi8 also have a chance to be satisfied with quail_kpi2.

Quali_kpi8 is consequent in the first rule form, which can be used to determine which another measurement (quali_kpi2 in this example) should be associated with it, in order to have a high level of satisfaction. A support of 74% for quail_kpi2= ok in the first association rule means that 74% of all the transactions under analysis shows that patients are satisfied or very satisfied with the interest and attention brought by medical staff.

For example for Rule 12 (Id=2) in Weka result, the confidence (conf) that the conditional probability that a transaction having quanti_kpi17 (The waiting time before registration and payment) is not acceptable by 14 patients, also contains quali_kpi11 (installation in the waiting room before registration) which is also not acceptable to 13 patients is 0.93. With this comparative study between the two results of the output of weka tool and that of our framework, it is easy to decide that Apriori algorithm functioned correctly.

After selecting the right key performance indicators, the domain expert analyzes the different results obtained from association rules. Those rules can be used to see what other KPIs should be taken into account to promote a high satisfaction with the selected KPIs. This analysis of frequent KPIs aims to find all interesting rules that correlate the presence of one set of KPIs with another set of KPIs. The results show that data mining technologies can be effectively applied to exploit semantic data and provide answers to users' queries. The originality of our work concerns how to explore and analyze frequent items set mining in order to search for recurring relationships in a given data set and then all found results are used for the discovery of interesting relationships between KPIs established in our ontology. The discovery of interesting correlation relationships among huge amounts of KPI transaction records can help in many businesses decision-making processes such as improving satisfaction and discover BP blockiness.



antecedent	support	consequent	support	conf
[quali_kpi2=null]	7	[quali_kpi8=null]	7	1.0
[quali_kpi2=ok]	74	[quali_kpi8=ok]	66	0.8918918918918919
[quali_kpi8=ok]	75	[quali_kpi2=ok]	66	0.88
[quali_kpi8=null]	11	[quali_kpi2=null]	7	0.6363636363636364
[quali_kpi8=Not ok]	14	[quali_kpi2=ok]	8	0.5714285714285714
[quali_kpi2=Not ok]	19	[quali_kpi8=ok]	9	0.473684210526315...
[quali_kpi8=Not ok]	14	[quali_kpi2=Not ok]	6	0.428571428571428...
[quali_kpi8=null]	11	[quali_kpi2=Not ok]	4	0.363636363636363...
[quali_kpi2=Not ok]	19	[quali_kpi8=Not ok]	6	0.3157894736842105
[quali_kpi2=Not ok]	19	[quali_kpi8=null]	4	0.210526315789473...
[quali_kpi8=ok]	75	[quali_kpi2=Not ok]	9	0.12
[quali_kpi2=ok]	74	[quali_kpi8=Not ok]	8	0.108108108108108...

Figure 3. Example 1of association rules using our framework

antecedent	support	consequent	support	conf
[quali_kpi11=null]	20	[quanti_kpi17=null]	19	0.95
[quanti_kpi17=Not ok]	14	[quali_kpi11=Not ok]	13	0.9285714285714286
[quanti_kpi17=ok]	36	[quali_kpi11=ok]	27	0.75
[quali_kpi11=ok]	46	[quanti_kpi17=ok]	27	0.5869565217391305
[quali_kpi11=ok]	46	[quanti_kpi17=null]	18	0.391304347826087
[quali_kpi11=Not ok]	34	[quanti_kpi17=null]	13	0.382352941176470...
[quali_kpi11=Not ok]	34	[quanti_kpi17=Not ok]	13	0.382352941176470...
[quanti_kpi17=null]	50	[quali_kpi11=null]	19	0.38
[quanti_kpi17=null]	50	[quali_kpi11=ok]	18	0.36
[quanti_kpi17=null]	50	[quali_kpi11=Not ok]	13	0.26
[quali_kpi11=Not ok]	34	[quanti_kpi17=ok]	8	0.235294117647058...
[quanti_kpi17=ok]	36	[quali_kpi11=Not ok]	8	0.2222222222222222
[quanti_kpi17=Not ok]	14	[quali_kpi11=ok]	1	0.071428571428571...
[quali_kpi11=null]	20	[quanti_kpi17=ok]	1	0.05
[quanti_kpi17=ok]	36	[quali_kpi11=null]	1	0.027777777777777...
[quali_kpi11=ok]	46	[quanti_kpi17=Not ok]	1	0.021739130434782...

Figure 4. Example 2 of association rules using our framework

```
Apriori
=======

Minimum support: 0.1 (10 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 1

Best rules found:

 1. quali_kpi2=ok 74 ==> quali_kpi8=ok 66    <conf:(0.89)> lift:(1.19) lev:(0.11) [10] conv:(2.06)
 2. quali_kpi8=ok 75 ==> quali_kpi2=ok 66    <conf:(0.88)> lift:(1.19) lev:(0.11) [10] conv:(1.95)
```

Figure 5. Example 1 of association rules using weka tools

```
Instances:     100
Attributes:    2
               quali_kpi11
               quanti_kpi17
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.1 (10 instances)
Minimum metric <confidence>: 0.1
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 5

Best rules found:

  1. quali_kpi11=null 20 ==> quanti_kpi17=null 19    <conf:(0.95)> lift:(1.9) lev:(0.09) [9] conv:(5)
  2. quanti_kpi17=Not ok 14 ==> quali_kpi11=Not ok 13    <conf:(0.93)> lift:(2.73) lev:(0.08) [8] conv:(4.62)
  3. quanti_kpi17=ok 36 ==> quali_kpi11=ok 27    <conf:(0.75)> lift:(1.63) lev:(0.1) [10] conv:(1.94)
  4. quali_kpi11=ok 46 ==> quanti_kpi17=ok 27    <conf:(0.59)> lift:(1.63) lev:(0.1) [10] conv:(1.47)
  5. quali_kpi11=ok 46 ==> quanti_kpi17=null 18    <conf:(0.39)> lift:(0.78) lev:(-0.05) [-5] conv:(0.79)
  6. quali_kpi11=Not ok 34 ==> quanti_kpi17=null 13    <conf:(0.38)> lift:(0.76) lev:(-0.04) [-4] conv:(0.77)
  7. quali_kpi11=Not ok 34 ==> quanti_kpi17=Not ok 13    <conf:(0.38)> lift:(2.73) lev:(0.08) [8] conv:(1.33)
  8. quanti_kpi17=null 50 ==> quali_kpi11=null 19    <conf:(0.38)> lift:(1.9) lev:(0.09) [9] conv:(1.25)
  9. quanti_kpi17=null 50 ==> quali_kpi11=ok 18    <conf:(0.36)> lift:(0.78) lev:(-0.05) [-5] conv:(0.82)
 10. quanti_kpi17=null 50 ==> quali_kpi11=Not ok 13    <conf:(0.26)> lift:(0.76) lev:(-0.04) [-4] conv:(0.87)
```

Figure 6. Example 2 of association rules using weka tools

V. CONCLUSION

The BP is rich with information to evaluate the performance of the Healthcare organizations and data mining has become a necessity. Performance measures enable ED to carry out a diagnosis of strengths and weaknesses of BP, which play a central role in the business process. Data mining can enable healthcare organizations to evaluate patient's satisfaction with the data value recorded from both quantitative and qualitative measurement by analysis of the frequent items and by making connections between KPIs related information. The raw data from KPIs are voluminous and heterogeneous. It needs to be collected and stored in a structured way. In this paper; we explored several KPIs using jBPM management features. The research provided a case study for the empirical application of healthcare process. Moreover, we analyzed some association rules related to estimated value instances in the healthcare BP. The contribution of this work is that KPI knowledge; understanding with patient experience in ED and interactions with other indicators are meaningful properties of BP improvement, for which our research questions are dedicated. The main benefit of using data mining knowledge discovery offers the opportunity to continuously improve the BP by manipulating relationships and later to identify the reason of bottlenecks. The final lists of interesting KPIs association rules force managers to rely on their intuition in order to manage potential candidate KPIs. In parallel, we plan to extend our framework based on our data mining association rules results to build a semantic presentation. For instance, we will have to delete or to discover new links between KPIs. This future planning will enable us to look for other opportunities of improvement for the BP and to prove the adequacy of our indicators and their relationships.

REFERENCES

[1] Kaskinen, Jani. "Creating a best-in-class KPI program." Strategic Finance 89.4 (2007): 29.

[2] V. Andrikopoulos, S. Benbernou, M. Bitsaki, O. Danylevych, M. Hacid, W. van den Heuvel, D. Karastoyanova, B. Kratz, F. Leymann, M. Mancioppi, K. Mokhtari, C. Nikolaou, M. Papazoglou, and B. Wetzstein. Survey on Business Process Management., July 2008.

[3] D. Parmenter. Key performance indicators: developing, implementing, and using winning KPIs.Wiley, 2009.

[4] Dumas, Marlon, et al. "Introduction to business process management." Fundamentals of Business Process Management. Springer Berlin Heidelberg, 2013

[5] A. Kronz. Managing of process key performance indicators as part of the aris methodology. In Corporate Performance Management: Aris in Practice, pages 31–44. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-30787-7.

[6] J. v. Brocke and M. Rosemann, "Handbook on business process management 1: Introduction, methods, and information systems,2014.

[7] G. Pitzos; M. Matsas and G. Chryssolouris. "Defining manufacturing performance indicators using semantic ontology representation", in CIRP 3, 2012.

[8] S. Opoku-Anokye and Y. Tang. "The design of a semantic-oriented organizational performance measurement system", in 14th Internacional Conference on Informatics and Semiotics in Organisation (ICISO), Stockholm, Sweden, p. 45-49. 2013. Available from http://centaur.reading.ac .uk/31975 [Accessed 10th March 2014].

[9] Y. Shen; D. Ruan and E. Hermans. "Modeling qualitative data in data envelopment analysis for composite indicators". International Journal of System Assurance Engineering, v. 2, n. 1, 2011.

[10] Claudia Diamantini, Domenico Potena, and Emanuele Storti. Extended drill-down operator: Digging into the structure of performance indicators. Concurrency and Computation: Practice and Experience, 2015.

[11] Ortega DEL-RÍO-ORTEGA, Adela, RESINAS, Manuel, DURÁN, Amador, et al. Using templates and linguistic patterns to define process performance indicators. Enterprise Information Systems, 2016, vol. 10, no 2, p. 159-192.

[12] Pedrinaci, C., Lambert, D., Wetzstein, B., van Lessen, T., Cekov, L., Dimitrov, M.: Sentinel: a semantic business process monitoring tool. In: Proceedings of the First International Workshop on Ontology-supported Business Intelligence (OBI). (2008)

[13] Wetzstein, B., Ma, Z., Leymann, F.: Towards Measuring Key Performance Indicators of Semantic Business Processes. In:11th International Conference on Business Information Systems (BIS'08). Innsbruck, Austria (2008) 227–238

[14] Claudia Diamantini, Domenico Potena, and Emanuele Storti. Sempi: a semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators. Future Generation Computer Systems, 54:352{365, 2016.

[15] Ahmad, Parvez, Saqib Qamar, and Syed Qasim Afser Rizvi. "Techniques of data mining in healthcare: a review." International Journal of Computer Applications 120.15, 2015.

[16] M. Patil, R. C. Joshi, and D. Toshniwal, "Association rule for classification of type -2 diabetic patients", Second International Conference on Machine Learning and Computing,2010.

[17] M. Ilayaraja and T. Meyyappan, "Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, 2013.

[18] Eapen, A. G. (2004). Application of Data mining in Medical Applications. Ontario, Canada: University of Waterloo, 2004.

[19] Eiko Kai et al., "Empowering the healthcare worker using the Portable Health Clinic", IEEE Transactions, DOI 10.1109/AINA.2014.108, 2014

[20] Usha Yadav, Gagandeep Singh Narula, Neelam Duhan, Vishal Jain and B. K.Murthy1, Development and Visualization of Domain Specific Ontology using Protege, Indian Journal of Science and Technology, Vol 9(16), April 2016.

[21] Simone Fiorini, Arun V Gopalakrishnan, Mastering jBPM6, Packt Publishing, 25 June 2015.

[22] Obe, Regina O., and Leo S. Hsu. PostgreSQL: Up and Running: A Practical Introduction to the Advanced Open Source Database. " O'Reilly Media, Inc.", 2014.

[23] Rebuge, Álvaro, and Diogo R. Ferreira. "Business process analysis in healthcare environments: A methodology based on process mining." Information systems 37.2, 2012.

[24] Ahmad, Parvez, Saqib Qamar, and Syed Qasim Afser Rizvi. "Techniques of data mining in healthcare: a review." International Journal of Computer Applications 120.15, 2015.

[25] Alpaydin, Ethem. Introduction to machine learning. MIT press, 2014.

[26] YILMAZ, Nergis et ALPTEKIN, Gülfem. An ontology-based data mining approach for strategic marketing decisions. 2013.

[27] PERAL, Jesús, MATÉ, Alejandro, et MARCO, Manuel. Application of Data Mining techniques to identify relevant Key Performance Indicators. Computer Standards & Interfaces, 2017, vol. 50, p. 55-64

[28] Ortega, A. On the Definition and Analysis of Process Performance Indicators. Diss. Ph.D. Thesis, University of Sevilla, Spain, 2012