IMPROVING HEALTHCARE INDUSTRY USING MINING TECHNIQUES

Maha ZAYOUD[1], Sorin IONESCU[2]

Healthcare is an important sector nowadays; it is related to many entities which needs the professionals to meet the different requirements. However, there are many issues appears in healthcare domain that attract the researchers to use different methods and techniques such as process mining, which is the practice of extracting knowledge in a specific organization and provides an accurate image of any system and suggests solutions to develop it. This paper defines the healthcare in general with its main needs, defines the main process mining techniques, and highlights the β algorithm and its impact on improving the healthcare from many perspectives.

Keywords: Healthcare, Business Process, Economics.

1. Introduction

Industry is the neural system in any country, and the growth of the industrial domains make the technologies, and the methods that can handle this growth is mandatory. However, the industrial domains are not limited to companies and organizations only, but they mean also the service sectors as well, such as the healthcare service sector. Healthcare is an important field of study because if its impact on the whole society and politics. Healthcare includes many aspects to study, and it is an attractive domain for researchers to find new methods of continues improvements.

The process mining is an interesting field of study for both researchers and practitioners, and many algorithms and methods are invented due to the importance of this field of study [1], but many methods are still facing problems when they are implemented in real life applications due to the complexity of some systems and their information management. There are many challenges and issues can be resolved by using one of the invented process mining techniques, such as the challenge of cost management in any domain, customer satisfactions, staff satisfactions, and finally the continues performance improvements. Using any of the process mining algorithms or methods requires having management information system that can store data of that organization, hence the process

[1] College of Engineering and Technology, American University of the Middle East. Kuwait, email: maha.zayoud@ aum.edu.kw

[2] Industrial Engineering Department, University POLITEHNICA of Bucharest, Romania, e-mail: sorin.ionescu@upb.ro

mining methods can discover the existed process and study its limitations and strengths. As a challenge of using the process mining methods is choosing the right method among all the existed methods to discover knowledge that can provide the organization with the required steps toward improvement [2]. However, the process mining techniques aim to develop, and discover the process based on the information extracted from event logs [4]. It is a result of the need of emerging between the data mining [3] and business process management [3], where data mining concerns with analyzing the different types of the data sets and the business process management focuses on modeling these data types and files. Hence, the process mining plays the middleware role between the two, to combine the data analysis with modeling. Moreover, process mining can handle the raw data or the event logs of any organization. There is no accepted benchmark to evaluate and compare the different proposed process mining algorithms which makes it difficult to select a suitable process mining algorithm for a given enterprise or application domain [4]. Process mining algorithms are used to mine business process models using process logs. The mined models are compared against the current process models of the enterprise for conformance checking that is a basic step of process mining. The next step, is to discover more efficient, streamlined business process models [2].

2. Healthcare System

The structure of healthcare system consists of medical care services under three different parties: primary care, hospitals, and community healthcare [5]. Each type aims to provide a specific kind of medical care, and each one is offered by the specialists' doctors, or any other staff who is responsible for this type of care service. The goal of the healthcare system is that good healthcare should be available to all people regardless of their financial situation. Providing preventive health services is the major function of the community health service [5]. Funding the healthcare sectors is one of the challenges that helps to ensure the quality services. The community health service refers to the cooperation with the local government in order to enable health and personal public care [6]. However, the healthcare system is funded from the general taxation system and the national insurance contributions [6]. These funds are divided among the health authorities to cover the medical expenses of each area. Although, the healthcare system is improved successfully, but it suffers from several problems such as shortage of resources, very demanding hospitals, waiting lists for low acuity patients and bad management. In the other hand, the advantages are worth to be highlighted, such as following the international standards for healthcare, and the public services are non-expensive service which grasp the attention of majority of individuals. Resources are directed by the system toward different areas and several types of

care based on the region needs. In parallel, primary care and hospital treatment is also covered by the private medical sector. Most private care are part of private medical insurance program. Healthcare is not an easy domain to study or develop because of the wide range of entities that consists of, or the different services that takes care of them, such as the medications parts, service cost, staff satisfactions, resources, researches, and data. As an example of one part of healthcare system is, hospital which is one major and top-level healthcare system, and can be divided into different departments. These departments are interacting with each other and sharing a variety of professional resources (Doctors, Nurses, Staff, etc.). This crossover leads to a very complex and sensitive system as mentioned because of the different departments and units in a hospital [6].

3. Process Mining Techniques

3.1. Business Process:

As seen previously that process mining is representing the middle step between the data mining, and the business process management. Data mining is defined as the process of analyzing data from different perspectives and summarizing it into useful information that can be used to improve many economic factors. While, the business process is defined as a collection of interrelated tasks, performed to achieve a business outcome which is a chain of tasks from purchasing to manufacturing to selling and delivering. Business process can be divided into two kind operational which is related to the core business and management which include the information system and strategic decisions [7]. Process innovation is the implementation of a new method for production by developing new products or services or changing the business model, or by developing new method of delivery. The innovation in any business can be achieved by changing the technology and environment of any origination. The advantages behind the process innovation like improving quality, faster processes, reduced labor costs, reduced materials, energy consumption and environmental damage.

3.2.   Process Mining Algorithms:

The process mining is the knowledge extracted from the event logs that is recorded by an information system applied in the organization [3]. Although, the information systems help to get the event logs, but these logs are rarely used to analyze the underlying processes. Hence, the mining techniques aim to develop, and discover the process based on the information extracted from event logs [3].It is a result of the need of emerging between the data mining and business process management, where data mining concerns with large data sets and the business process management focuses on modeling these data. Hence, the process mining

plays an important role between the two, to combine the data analysis with modeling. Moreover, process mining can handle the raw data or the event logs of any organization. However, process mining algorithms are mainly categorized as either local algorithms such as the α algorithm [8] and the heuristic [9], or global algorithms such as genetic [10] and fuzzy algorithm [11].

Local algorithms are based on local information about events in a log [12], which is used to detect the dependencies between these events by extracting information about what tasks directly precede or directly follow each other in that log. The global algorithms are defined as non-local constructs that cannot be determined by only looking at the direct successors and predecessors which is the local context of a task in a log [12].

The process mining algorithms and techniques face many challenges such as:

- Incomplete and noisy input data.

- Distinguishing sequences, forks, and concurrency.

- Deriving block-structured and arbitrary loops.

- Distinguishing repeated activities especially in the loops.

- Dealing with fuzzy process entry and end points.

- Detecting different process types and variants.

Managing a critical system such as the healthcare system requires a suitable algorithm which can manage the above challenges to give an accurate process model. The well know algorithms of process mining have many advantages and disadvantages, and many of them have extension like α algorithm that is simple and easy to use but it is not suitable for real life application, has an α miner for example. Because of these challenges, the need to use a reliable algorithm is needed especially in out domain which the healthcare system, therefore the β algorithm is used in this study and a brief explanation about it is mentioned in the next sections.

4. The β Algorithm

The β algorithm framework [12] is a process mining algorithm that is designed to overcome many issues appear in other process mining methods. The β algorithm is a learning technique to extract knowledge about any system by learning different log files of that system [12]. This framework is a probabilistic and it learns the process in an accumulative manner. The learning process is divided into phases as shown in the flowchart of Figure 1. First, it starts with learning the set of events in the log files then, it calculates the dependencies among those events. After learning the dependencies, the probability distributions of those events and

their correlation are detected [12]. A knowledge base is integrated to the learning framework to predict the coming events out of the learned model. The knowledge base has few simple rules for real-time prediction of events. The β algorithms starts with analyzing the data files first, where all required information about that system is input to the framework, then the knowledge discovery starts, with the learning of the events, dependencies, probabilities, and frequencies all will build the work flow model to reach at the end what is called a predict model which describes the current process clearly. The β algorithm is not a theoretical solution only; it is proved mathematically, which is the accepted method of proving any new technique [12]. Then, it is written in an algorithmic format and it requires data in terms of events (activities of all resources in the system), where each event should have a starting time and ending time to deploy the rules of this framework correctly [12].
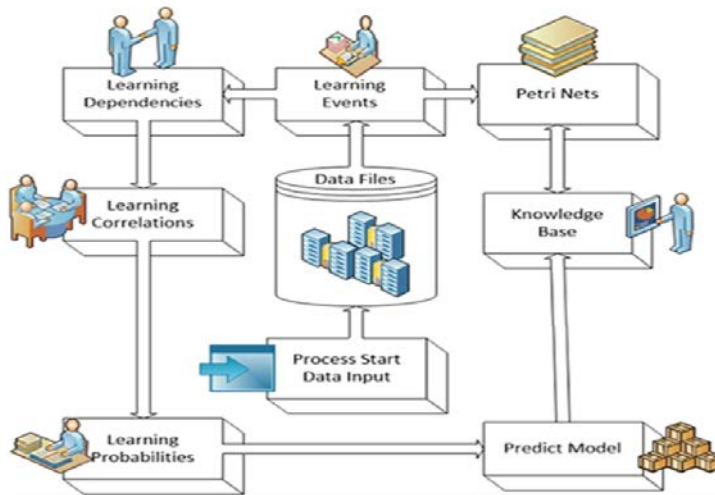

Fig 1. The β algorithm Framework Flowchart [12].

5. Applying the β Algorithm Framework on Healthcare System

This section shows how the process mining methods is applied on real life application. The example is data input represents a log file of patients' activities in a hospital. The β algorithm is chosen among other methods due to its many advantages that are presented in Fig. 2.

Item	β	α	Fuzzy	Genetic	Heuristic
Noisy input data:	√	×	√	√	√
Concurrency:	√	√	√	√	√
Loops:	√	×	×	√	×
Repeated Squences :	√	×	×	√	×
Fuzzy process :	√	×	√	×	√
Variants detection :	×	×	√	×	×

Fig 2. Famous Process Mining Algorithms VS. Challenges [12].

Where $\checkmark$ means this algorithm can handle the required challenge, and X means it cannot handle it. As seen in Figure 2, the β algorithm can manage many challenges that face any method to detect the process and all its aspects. This framework is chosen due to its advantages in detecting the events in any system after learning the log file. Then, the β framework learns how these events are related to other, such as which event follow another event, and the different logical relations that may appear between events are also detected which helps to build the knowledge base toward the existed process. The β framework is also a probabilistic approach as shown in Figure 1, hence it suits the real-life applications more than other techniques to detect high risk factors and the disadvantages of the current process [12]. The needs of using process mining techniques, especially like the β algorithm which is suitable for healthcare system and can be applied on real data, because building a process from scratch is easier than trying to detect the current process of an existed system and knowing all its aspects accurately, with all advantages and disadvantages which opens the way for improvement and development of an organization without causing huge loss of data or resources.

Item	β	α	Fuzzy	Genetic	Heuristic
AND operation: ∧	√	√	×	×	×
OR operation: ∨	√	√	×	×	×
XOR operation: ⊕	√	×	×	×	×
Implication : →	√	√	×	×	×
Partial Parallelism : ↔	√	√	×	×	×
Full Parallelism : ⇔	√	×	×	×	×
Probabilistic	√	×	×	×	×
Local	√	√	×	×	√
Global	×	×	√	√	×

Fig 3. Events' relations versus Process Mining Algorithms [12]

5.1.   The Economic Benefits of Applying the β Framework in Healthcare

There are many advantages of using process mining methods, such as providing knowledge about the existed process in the system, and how this process works based on evidence and using the log data files that are stored in the management information system of an organization, which allows an objective reconstruction of the process flows [13].

The benefits of using Process mining algorithms are:

1. Time factor:

Discovering the process activities needs several weeks or months, which is a considerable time in any system, implementing a process mining can come up

with first results and hypotheses quickly, which can help to increase trust and engage people with the services that are provided in that hospital without consuming a considerable time using the traditional methods [13].

2. Understanding the system:

Process mining provides an objective reference on how things are done and giving the reason of why people work the way they do, which is rarely detected using the observation or data analysis. However, understanding the root causes of inefficiencies also on the human level is crucial to successfully implement organizational change, hence it maximizes the value by getting deep knowledge that is normally not discovered [1].

3. Get a head start within new domains

Process mining can help consultants, who are specialists in specific domain, and able to provide assistant in their domain only, to understand the new domains also and approach their job by having a tool to understand the process quickly, and increase the productivity especially of junior analysts, right from the start [13].

4. Help Clients to justify Changes within the Company

Process mining can provide an objective reference for the clients of any organization, which can put them in a stronger position to achieve their goals successfully [13], which getting the right care in the case of the hospital.

5. Compare "before" and "after"

As mentioned before, the process mining can understand the process quickly, especially for process improvement projects, where we need to demonstrate the effect of the implemented changes. For example, showing how the process has been streamlined after a change (and one month of new data collection) by comparing the "before" and "after" images. Ideally, the process performs much better now, and you can use process mining to communicate these results [13].

Hence, the above benefits show the impact of using the process mining techniques on any organization especially by reducing the time of analysis, the cost of understanding the process, and getting the objective references of clients' goals and here are the patients and the medical staff as well.

6. The β algorithm on hospital example

The β algorithm is implemented on a hospital that has some issues related to time consumption in serving patients, cost, staff satisfactions and patient's satisfactions as well. The following is the case study of that hospital. The data input to the β platform is many log files that include one patient's activities with their timing

information, and many patients' events as well. Figure 4. And Figure 5 Show examples about what are the basic events for a patient in that hospital.

7:45	Arrival	5
7:53	Checking Urgent	15
8:10	Waiting room	35
8:59	Doctor Decide	25
9:30	Examination	60
10:45	Treatment	50
11:40	Release resource	5
12:00	Billing	20

Fig4. One Patient Log File [12]

Time	Event	Duration
9:05	Arrival	10
9:06	Data Collection	5
9:15	Waiting room	70
9:59	Doctor Decide	20
9:30	Radiology	40
10:45	Treatment	40
11:40	Release resource	5
12:00	Billing	20

Fig 5. Another Patient Log File [12]

The patient's information that is categorized as events such as arriving to hospital, checking urgent, billing, etc. all input to the platform to be processed and studied. Then the dependency matrix between these events is an output result of this information such as Figure 6. Where the 1 means their dependency between this event located in the row and the other event located in the column of the matrix, 0 means there is no dependency between these two events. As an example, the billing event depends on discharging the patient after getting the required treatment by the doctor.

$$\mathbb{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \qquad \mathbb{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Fig 6. Dependency Matrix of First Log File   Fig 7. Dependency Matrix of First Log File [12].

The β platform can know the relation between the events of patients, such as which event can start with which event, and which event cannot start at the time the other one is executed. Moreover, the frequency of the events and their

relations is calculated using the platform, hence the probabilities of these events are detected. In fact, the β algorithm [12] is an extension of the α algorithm [8], which is originally builds a workflow nets out of the events log file and can use Petri nets [14]. Hence, the Petri net that is discovered from the output results of the β platform is presented in Fig. 8.
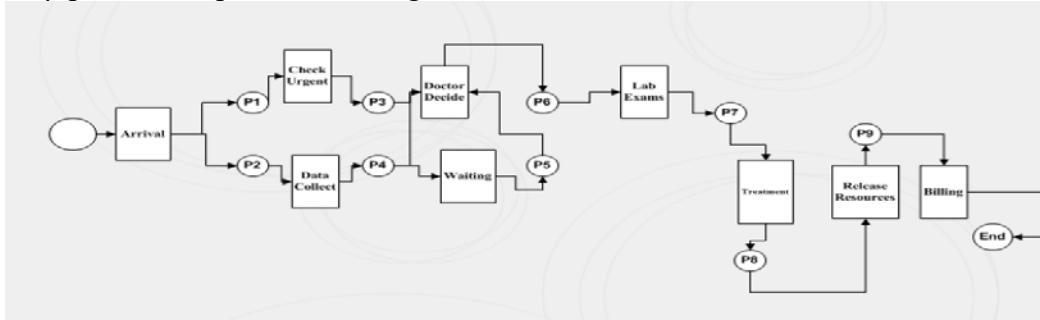

Fig 8. Petri Net for Patients Events in the Hospital.

After that, and from detecting all the required information and events and their relations and probabilities, the process model is discovered as shown in Figure 9. This process model is a starting point to understand the current system with all its advantages and disadvantages, such as the deadlock in the system.
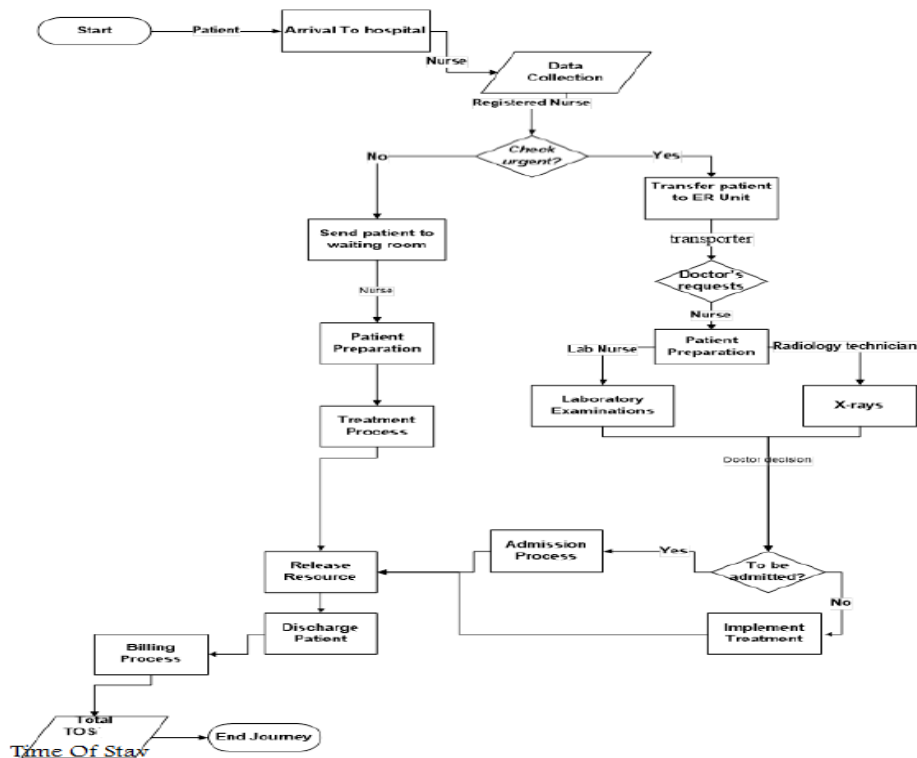

Fig 9. A Process Model for a Hospital [12].

7.   Results Analysis of the Process Mining

Analysing the results means studying the strengths, weaknesses, opportunities and threats, identifying and analyzing the internal and external factors that can have an impact on the viability of an organization, product, place or person [15][17]. Business entities are the main users of the analysis, the analysis is used to show the benefits on the industry of using the process mining techniques, since it shows the different factors that needs to be analysed. Using a huge amount of patient's information that includes the patient's activities and their starting and ending time with the duration, shows the number of detected events and their probabilities which leads to know more about the system and its events [16], hence the deadlock can be detected and suggestion on better solution can be implemented. Fig. 10. Shows how many events are detected from a log file of big number of patients. In Fig. 11. We can notice how the probabilities of each event in the hospital system is detected.
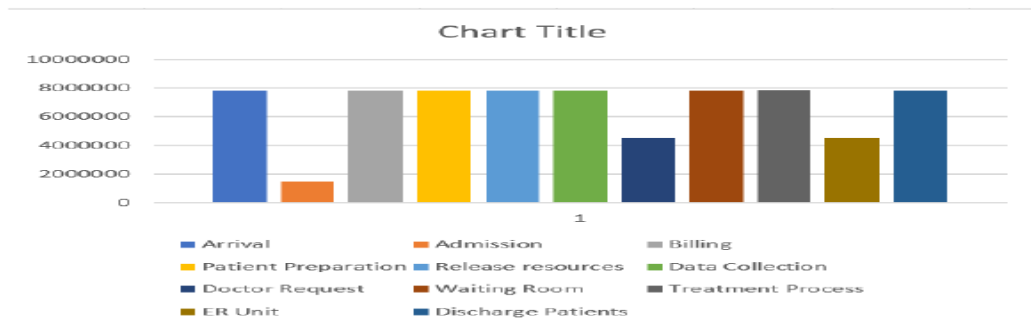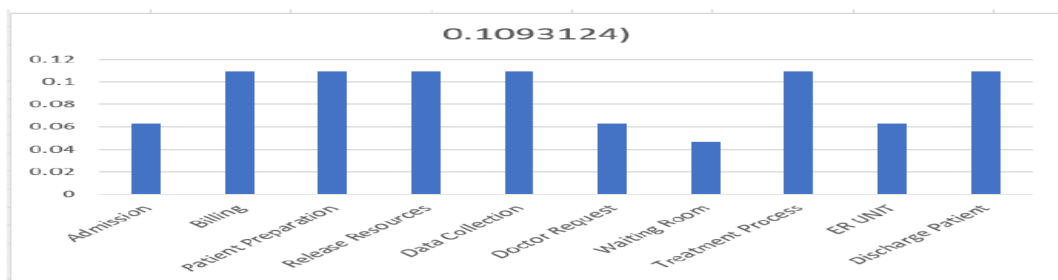

Fig 10. Events Statistics

Fig 11. Probability of Events

8.   Conclusion

As a summary, this paper presents the benefits of using the process mining techniques to improve the operational tasks in an organization. It is also explains

the challenges of using process mining algorithms, such as ease of use, match with the organization needs, and improving the current process model. However, this paper focusses on healthcare industry and mentions its main issues such as long time of stay in a hospital for patients, high costs of services, and lack of resources and data. In this research, the β algorithm framework which is an extension of the well-known algorithm named α algorithm, is implemented on hospital's example because the hospital is a top entity in the healthcare system. The choice of the β algorithm among other known process mining methods is done because this algorithm is proved to be valid in healthcare domain, and it is able to extract an accurate knowledge for log files of hospital. This knowledge provides the actions in that particular hospital and the relations between those events, all will help to build a process model that is able to highlights the strengthens and weaknesses for this system, hence to suggest solution for some problems. However, choosing the right mining method for an organization is a challenge, hence choosing the β algorithm went through many phases and decisions making. After that, the log files information from a particular hospital is input to the β platform, and the output results are simulated to build a knowledge base about the system of that particular hospital. The choice of using the β algorithm is also proved based on economic factors related to the cost of services, patients and management satisfactions from using the process mining techniques, and they mentioned in this paper. Finally, the process mining is important to be implemented in the industrial domains that have issues and struggle to maintain their business and build continuous success, in addition to manage big number of clients, all those facts, lead those organizations and systems to improve their processes, increase the overall performance, and achieve the required goals of those systems. All the mentioned facts and improvements cannot be achieved without choosing the suitable method for an organization.

R E F E R E N C E S

[1] *Van Der Aalst*, Wil. Process mining: discovery, conformance and enhancement of business processes. Vol. 2. Heidelberg: Springer, 2011.

[2] *Van Der Aalst, Wil, Arya Adriansyah, Ana Karla Alves De Medeiros, Franco Arcieri*, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose et al. "Process mining manifesto." In International Conference on Business Process Management, pp. 169-194. Springer, Berlin, Heidelberg, 2011. [3] Laudon, Kenneth C., and Jane P. Laudon. Management information systems. Prentice Hall PTR, 1999.

[4] *Van Der Aalst*, Wil. Process mining: discovery, conformance and enhancement of business processes. Vol. 2. Heidelberg: Springer, 2011.

[5] *Andersen, Ronald M., Thomas H. Rice, and Gerald F. Kominski.* Changing the US health care system: Key issues in health services policy and management. John Wiley & Sons, 2011.

[6] *Chaulagai, Chet N., Christon M. Moyo, Jaap Koot, Humphrey BM Moyo, Thokozani C. Sambakunsi, Ferdinand M. Khunga, and Patrick D. Naphini.* "Design and implementation of a health management information system in Malawi: issues, innovations and results." Health policy and planning 20, no. 6 (2005): 375-384.

[7] *A.J.M.M. Weijters, W.M.P. van der Aalst, and A.K. Alves de Medeiros,*" Process Mining with the Heuristics Miner Algorithm".

[8] *Van der Aalst, Wil, Ton Weijters, and Laura Maruster.* "Workflow mining: Discovering process models from event logs." IEEE Transactions on Knowledge and Data Engineering 16, no. 9 (2004): 1128-1142.

[9] *Weijters, A. J. M. M., Wil MP van Der Aalst, and AK Alves De Medeiros.* "Process mining with the heuristics miner-algorithm." Technische Universiteit Eindhoven, Tech. Rep. WP 166 (2006): 1-34.

[10] *De Medeiros, AK Alves, and A. J. M. M. Weijters.* "Genetic process mining." In Applications and Theory of Petri Nets 2005, Volume 3536 of Lecture Notes in Computer Science. 2005.

[11] *Günther, Christian W., and Wil MP Van Der Aalst.* "Fuzzy mining–adaptive process simplification based on multi-perspective metrics." In International conference on business process management, pp. 328-343. Springer, Berlin, Heidelberg, 2007.

[12] *M. Zayoud, Y. Kotb, S. Ionescu,* β Algorithm: A New Probabilistic Process Learning Approach for Big Data In Healthcare. IEEE Access Journal PP (99):1-1 June 2019 DOI: 10.1109/ACCESS.2019.2922635.

[13] https://fluxicon.com/blog/2010/05/7-reasons-for-consultants-to-do-process-mining/.

[14] Reisig, Wolfgang. Petri nets: an introduction. Vol. 4. Springer Science & Business Media, 2012.

[15] *S. Asghar, and G. Vaidyanathan.* SWOT analysis and theory of constraint in information technology projects. Information systems education journal 2, no. 23 (2004): 3-19.

[16] *A. Puneeta,* Ranking the strategies for Indian medical tourism sector through the integration of SWOT analysis and TOPSIS method. International journal of health care quality assurance 30, no. 8 (2017): 668-679.

[17] *P. Kumar, and I.Rathore*, The need of mining industry–a SWOT analysis. International Research Journal of Earth Sciences 3, no. 8 (2015): 32-36.