

Received October 9, 2020, accepted November 10, 2020, date of publication November 18, 2020, date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039111

# A Paradigm-Shifting From Domain-Driven Data Mining Frameworks to Process-Based Domain-Driven Data Mining-Actionable Knowledge Discovery Framework

FAKEEHA FATIMA<sup>1</sup>, RAMZAN TALIB<sup>1</sup>, MUHAMMAD KASHIF HANIF<sup>1</sup>, AND MUHAMMAD AWAIS<sup>2</sup>

<sup>1</sup>Department of Computer Science, Government College University, Faisalabad 38000, Pakistan

<sup>2</sup>Department of Software Engineering, Government College University, Faisalabad 38000, Pakistan

Corresponding author: Ramzan Talib (ramzan.talib@gcuf.edu.pk)

**ABSTRACT** The success of data mining learned rules highly depends on its actionability: how useful it is to perform suitable actions in any real business environment. To improve rule actionability, different researchers have initially presented various Data Mining (DM) frameworks by focusing on different factors only from the business domain *dataset*. Afterward, different Domain-Driven Data Mining (D3M) frameworks were introduced by focusing on *domain knowledge* factors from the context of the overall business environment. Despite considering these several *dataset* factors and *domain knowledge* factors in different phases of their frameworks, the learned rules still lacked actionability. The objective of our research is to improve the learned rules' actionability. For this purpose, we have analyzed: (1) what overall actions or tasks are being performed in the overall business process, (2) in which sequence different tasks are being performed, (3) under what certain conditions these tasks are being performed, (4) by whom the tasks are being performed (5) what data is provided and produced in performing these tasks. We observed that the inclusion of rule learning factors only from *dataset* or from *domain knowledge* is not sufficient. Our Process-based Domain-Driven Data Mining-Actionable Knowledge Discovery (PD3M-AKD) framework explains its different phases to consider and include additional factors from *five perspectives* of the business process. This PD3M-AKD framework is also in line with the existing phases of current DM and D3M frameworks for considering and including *dataset* and *domain knowledge* accordingly. Finally, we evaluated and validated our case study results from different real-life scenarios from education, engineering, and business process domains at the end.

**INDEX TERMS** Actionable knowledge, business process, data mining, data mining framework, domain-driven data mining framework, data privacy.

## I. INTRODUCTION

In today's competitive and dynamic business environment, the discovery of actionable knowledge has become an exhaustive task as data mining frameworks were and are still concentrating on considering dataset factors, i.e., data types, dimensions, quality of data, etc. from dataset alone to take organizational business decisions [1]. The discovery of actionable knowledge highly depends upon the learned rules that can directly or explicitly determine the specific context that leads to actions [3]. Context refers to as "any information that can be used to characterize the situation of an

object" [4]. Context determines, "A system is context-aware if it uses contextual information or services to provide a relevant outcome to the users, where relevancy depends on user's tasks or actions" [5]. So, actionability determines how much these contextual learned rules are useful for making actionable decisions in a real business environment [6]. To make organizational decisions by concentrating and considering only the datasets through data mining (academia proposed) frameworks may silo mislead the actual representation owing to the absence of contextual part about the business domain according to which data was created actually.

As in traditional DM frameworks [7]–[10], data miner extracts the organizational data that are stored in different Information System (IS) to generate hidden interesting

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi<sup>1</sup>.

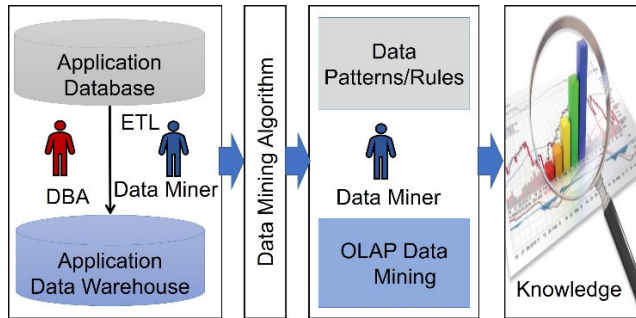


FIGURE 1. The traditional data mining process.

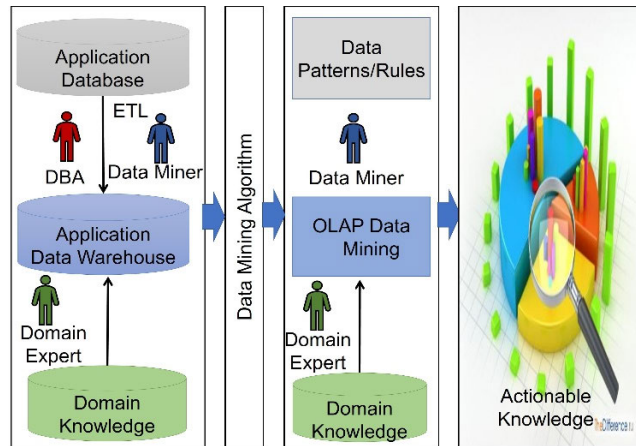


FIGURE 2. Domain-driven data mining for actionable knowledge discovery.

patterns and rules for the discovery of knowledge to take organizational decisions as exhibited in FIGURE 1. Therefore, the creation of edifice learned rules based on the automatic assortment from the dataset that may be ‘innovative’ and ‘novel’ but not actually being accurately fitted in the real-world business setting [11], as it lacks the involvement of concerned contextual factors and domain experts while generating the learned patterns. However, these factors alone are not adequate for taking decisions to attain knowledge, as did not provide the actual context that may cause the interestingness gap between generated rules and the business needs.

To overcome the issues, while generating actionable patterns for actionable knowledge discovery, D3M frameworks have been presented [3], [6], [10], [12]. The introduction of D3M frameworks [6] revolutionized the field of discovering actionable knowledge by transforming the “data-centric knowledge discovery process” to “domain-driven data-centric actionable knowledge discovery process,” according to the real-world business needs.

D3M tries to reduce this academia-industry gap by adding domain knowledge factors with experts experiences in the DM process, as shown in FIGURE 2 while generating patterns to make knowledge actionable [13]. Domain knowledge is a specialized ‘background knowledge’ related to user experiences, value, and insight required to attain mastery and accuracy for a concise and clear understanding of the

domain-specific sphere. Domain experts [7] are the persons who have the expertise to solve the specific problems in their respective fields. They gain expertise by doing similar tasks and solving problems according to the circumstances; after that, the method is stored, and a rule of thumb has been made to solve this type of situation [9]. The involvement of domain experts (i.e., data miner, domain experts) and domain knowledge factors such as background knowledge [9], environment [10], constraints [11], users, etc., helped to analyze the learned rules according to business context to discover actionable knowledge.

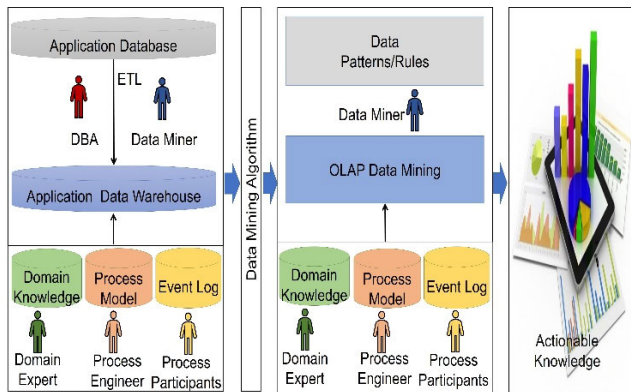
These learned patterns generated by D3M frameworks are although more closely associated with real-world business needs for discovering actionable knowledge than those formed out using traditional data mining, i.e., Knowledge Discovery Database (KDD) frameworks. Still, there is a gap between D3M generated learned rules and achieving actual business goals due to missing the process-centric context while making erudite learned rules for taking actionable decisions [14].

The objective of our research is to improve the learned rules’ actionability. Actionability is based on the interestingness of erudite patterns, must be the consequence of a good balance between the technical and business context. For this purpose, we have analyzed: (1) What overall actions or tasks are being performed in the business process. In a business environment, processes are the fundamental building blocks of organizational success. It played a crucial role compared to any other factor by determining the contextual information about the process which is executing its tasks in the real-world. Whether it is a government or semi-government, each organization has to manage numerous processes, i.e., order-to-cash, sales-to-purchase, etc. Process and its execution environment tell the underlying story to adapt to the new circumstances and complying with the fast-growing business requirements [15].

(2) In which sequence, different tasks are being performed. Typically, the business process has a detailed description prescribing about the context regarding how the underlying task must or should be executed, i.e., described the real world-way dependencies among tasks, as the output of one task becomes the input for another task.

Data that has been generated by these underlying tasks are recorded in the event log. An event log chronicles the underlying record of what is actually happening in the underlying information system. So, the conceptual accumulation of process generated data has emerged as a new way to gain a clear picture of the underlying task behavior. Deprived of adding an appropriate process-oriented execution context can’t work effectively because the outcome of one process task becomes the input of the next task [16].

(3) Under what certain conditions these tasks are being performed. Moreover, it helps to distinct the typical tasks into well-defined activities by identifying the roles and task dependencies in the business processes. Each task is assigned a set of triggering conditions and generates an effect of



**FIGURE 3. Process-based domain-driven data mining-actionable knowledge discovery process.**

its execution. Because the output of one task becomes the input for another task, these input and output information have been represented as data entities, defined the primitive or complicated data variables in the information system [17].

(4) By whom the tasks are being performed determines the role, i.e., agents who are suitable to execute a particular process to achieve a specified business goal. To accomplish specified tasks, the ‘role of actor’ played a crucial part.

(5) What data is provided and produced in performing these tasks? Data generated by these tasks cavorted a crucial role. As a process is nothing without data, and data is meaningless without the context and ignoring the process in which it accomplishes its tasks.

Data mining, domain-driven data mining, and process mining technologies are being used separately in their respective world. To eliminate these limitations and issues, there needs to be shifting the paradigm from “Domain-Driven Data Mining” towards “Process-based Domain-Driven Data Mining for Actionable Knowledge Discovery”. By consolidating data, domain, and process relevant factors, that leverage to bridge these technologies together as shown in FIGURE 3. Consequently, the decisions taken by entailing these process generated data according to domain knowledge, make erudite learned rules more actionable [18].

Therefore, our PD3M-AKD framework revolutionized the field of the mining process. It helps to incorporate these business process perspectives with domain and data factors in various phases of mining frameworks, to make learned patterns more actionable and interesting for enchanting smarter decisions that are extra closer to real-world business needs.

This paper gives a detailed overview of our PD3M-AKD framework that demonstrates real-life scenarios from educational, engineering, and business process domains to evaluate the results. Moreover, it empowers decision-makers to orchestrate domain experts, process executer, and database organizers regarding domain knowledge to make efficient and more effective actionable business decisions to achieve organizational goals.

The rest of the paper is structured as follows. Section 2 presents a review of existing DM and D3M frameworks and factors affecting actionable knowledge

discovery. Section 3 presents our PD3M-AKD framework inclusion of additional factors from five business process perspectives to find erudite learned rules and its usefulness in the real business environment to discover actionable knowledge. In Section 4, we evaluated and validated our case study results from different real-life scenarios from education, engineering, and business process domains at the end. Finally, In Section 5, concluding remarks and future directions have been demonstrated.

## II. RELATED WORK: EXISTING DM AND D3M FRAMEWORKS

This section aims to present the literature review related to different DM and D3M frameworks closely related to our proposed PD3M-AKD framework. We have reviewed many leading frameworks that are recurrently being used and play a leading role in the mining process from both domains. DM frameworks, i.e., KDD [11], [13], Cross-Industry Standard Process for Data Mining (CRISP-DM) [14], Data Mining Methodology for Engineering (DMME) [12] and D3M frameworks, i.e., Domain-Driven In-Depth Pattern Discovery (DDIP-PD) [3], Data Mining Integrated with Domain Knowledge (DMIWDK) [9], Loop-Closed Iterative Refinement Model [6], Domain-Driven Data Mining-Actionable Knowledge Discovery (D3M-AKD) [19], Domain-Driven Data Mining Knowledge-AKD (D3MK-AKD) [10].

In this paper, we have analyzed that the learned rules actionability and interestingness are based on how much conferring to these different DM factors (i.e., data context, mining goals, technical context), D3M factors (i.e., domain knowledge, domain experts, business context), and Process-based factors (i.e., process description, process sequence, process conditions, process performer and process data), for enchanting smarter decisions, that are extra closer to real-world business needs. Finally, we will present the summary table of our review to foster a better understanding of the PD3M-AKD framework.

KDD [11], [13] is an iterative and interactive process that comprises six phases (i.e., selecting the data, pre-processing, transformation, data mining, interpretation, and evaluation of patterns) consolidating the technical factors for the discovery of knowledge. In these phases, KDD had excellent support to focus on just DM factors: *data context*, goal mining, and technical *context* for generating more interesting and innovative patterns. Nevertheless, it doesn’t support the D3M factors, i.e., *domain knowledge*, *domain experts*, and *business context*, and Process-based factors, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data* that are affecting in the real business environment, while performing the mining task. Therefore, deprived of entailing these D3M and process-based factors lacks integrity and adeptness to depict the actual business needs while achieving actionable knowledge.

DMME [12] specifically designed for the fourth industrial revolution and engineering applications that provide a holistic view for analyzing analytical business decisions in the pro-

duction domain that support the satisfaction of necessary preconditions for successful implementation of data-driven processes analysis. It has very good support to transform the business goals into measurable technical goals by accumulating technical understanding, technical realization, and technical implementation regarding domain by extending the existing CRISP-DM framework. Nevertheless, still without the involvement of domain experts regarding business context and process-based factors, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data* in the mining process that is affecting in a real business environment.

CRISP-DM [14] is well-organized, structured, and precisely defined as an industrial framework to extract knowledge for business success according to business needs. It entails a cycle that encompasses six phases (i.e., business understanding, data understanding, data pre-preparation, modeling, evaluation, and deployment). Although, these phases have adequate support to different DM factors, i.e., *data context*, *mining goals*, and *technical context* with the *business context* have been considered during the mining process. However, deprived of concentrating D3M factors, i.e., *domain knowledge*, *domain experts*, and *Process-based factors*, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data*, the mining process do not depict the actual business situation. So, the generated learned rules may be more innovative and efficient, but difficult to adapt in the real business environment.

DDIP-PD [6] has been developed as an iteratively interactive in-depth pattern mining process in a constrained-based manner that embeds significant support in domain-specific context with domain experts involvement that eradicates the complexities of the knowledge discovery process. The DDIP-PD comprises numerous phases: 1. problem understanding, 2. constraint analysis, 3. data understanding, 4. data pre-processing, 5. modeling, 6. result evaluation, 7. results post-processing, 8. deployment, 9. knowledge delivery, and report synthesis for smart decision making. To enhance the actionability of learned rules, DDIP-PD phases consider the good support to the DM factors, i.e., *data context*, *mining goals*, *technical context*, and D3M factors, i.e., *domain knowledge*, *domain experts* (technical and business analysis), *business context*, to a degree it satisfied both the technical and business needs to achieve the maximum mining goals according to business context. Although, DDIP-PD incorporate numerous factors based on data and domain knowledge but still ignore the presence of process-based factors (i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data*) that are representing the real business context while generating learned rule for the discovery of actionable knowledge.

D3MK-AKD [9] considered the business's technical, economic, social aspects for developing and deploying actionable knowledge. In these phases (i.e., business understanding, ii. constraints analysis, iii. pre-processing

iv. modeling, v. Post-processing, in-depth mining patterns, vi) deployment, vii) feedback) good consideration of the DM factors, i.e., *data context*, *mining goals*, *technical context*, and D3M factors, i.e., *domain knowledge*, *business context* from the real world, and make assure that the data being restructured and correct according to business demand. However, to some extent, it supports the domain experts involvement during the mining process. However, while analyzing the D3MK-AKD framework, we have observed that although considered the DM and D3M factors, enhancing the technical and business interestingness helps to improve learned patterns' actionability. It still ignored the process-based factors, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data*, that depict the actual execution of tasks to generate erudite learned rules for making actionable business decisions.

In [10] examined DMIWDC to solve the real-world problems and attain actionable knowledge by integrating DM factors: i.e., *data context*, *mining goals*, *technical context*, and D3M factors, i.e., *domain knowledge*, *domain experts*, the *business context* that played a crucial role in a constraints-based manner. To find the learned rules actionability, domain knowledge has been imparting excellent support in different phases, i.e., understanding the mining tasks, pre-processing the data according to the domain experts experience, selecting the algorithms, and tuned the parameter values to find interesting and actionable knowledge. Although DMIWDC improved the knowledge discovery process by adding DM and D3M factors, it still did not consider process-based factors, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data* that depict the actual situation of the business process. Without considering these processes, associated factors didn't depict the underlying execution of tasks impacting their part while generating erudite actionable patterns.

In Loop-closed Iterative Refinement Model [11] presented the extended version of DDIP-PD for actionability enhancement in the real-world environments rather than demonstrating algorithms. The outcome of this retrospection and rethinking is a paradigm-shifting from traditional data-driven data mining towards domain-driven target-oriented research and development by considering the different DM factors (i.e., *data context*, *mining goals*, *technical context*), D3M factors (i.e., *domain knowledge*, *domain experts*, *business context*) the constraints-based manner in the knowledge discovery process for satisfying real business user's needs. Despite considering these several *datasets* and *domain knowledge* factors in different phases of their frameworks, the learned rules still showed a lack of actionability due to missing support of process-based factors, i.e., *process description*, *process sequence*, *process conditions*, *process performer*, and *process data*. There is needed to add the contextual process-based factors to take more actionable business decisions while generating erudite learned rules.

D3M-KDD [19] has specifically developed by considering the DM factors, i.e., *data context*, *mining goals*, and



TABLE 1. Review summary of DM, D3M, and Process-based factors.

Frameworks	DM factors			D3M factors							Process-based factors		
	Data Context	Technical Context	Mining Goals	Domain Knowledge	Domain Experts	Business Context	Process Description	Process Sequence	Process Conditions	Process Performer	Process Data		
DM	KDD	●	●	○	○	○	○	○	○	○	○		
	CRISP-DM	●	●	○	○	○	○	○	○	○	○		
	DMME	●	●	○	○	○	○	○	○	○	○		
	DDIP-PD	●	●	●	●	●	○	○	○	○	○		
D3M	D3M-AKD	●	●	●	●	●	○	○	○	○	○		
	D3MK-AKD	●	●	●	●	●	○	○	○	○	○		
	DMIWDK	●	●	●	●	●	○	○	○	○	○		
	Loop-closed Iterative Refinement Model	●	●	●	●	●	○	○	○	○	○		
		●	●	●	●	●	○	○	○	○	○		

Legends: ○ Don't Support ○ Partially Support ○ Averagely Support  
 ● Good Support ● Excellent Support

technical context, with the maximum consideration of D3M factors, i.e., domain knowledge, domain knowledge experts according to business context for taking interesting and actionable business decisions. The actionability of learned rules has a more significant influence on making decisions than just exploring the technically innovative and efficient rules. D3M-KDD has been entailed an effective and practical approach by involving technical and business context with domain experts in the AKD process. However, the absence of process-based factors, i.e., process description, process sequence, process conditions, process performer, and process data hindered the discovery of erudite learned rules for applicable, actionable knowledge discovery fulfilling business adeptness.

Table 1 presented the review summary of DM, D3M, and Process-based factors discussed in the literature.

### III. METHODOLOGY

The evaluation of the PD3M system caters to the significance of factors for generating erudite actionable patterns (P) from technical and business contexts in different phases of the PD3M-AKD framework. Different types of data, domain, and event log repositories have been used to extract data from numerous resources, as shown in FIGURE 4.

**Data Repository (DR):** It stored the operational data generated by the Information System maintained by the database administrator. DR contained Data Mining factors (DMf) (fi,..fn) that represent DMf (i.e., data context, mining goals, technical context, etc.) selected by the data miner to perform the mining operation [21].

#### Event-log

**Repository (ER):** It holds the real-world process generated data with their attributes such as the case-id, the timestamps, the resources, etc. relevant information in the event log [14]. The process engineer maintained the event log by attaining information from the process participants. The event log

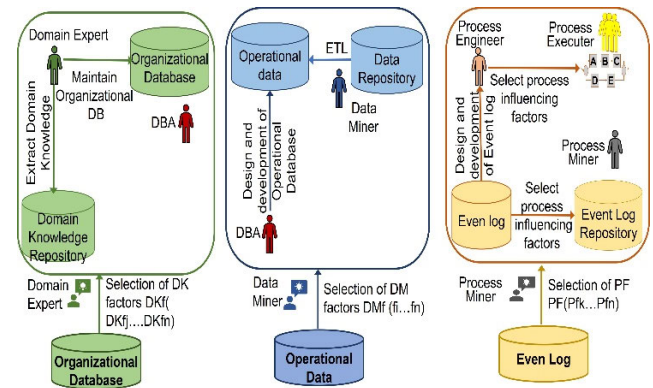


FIGURE 4. PD3M-AKD repositories.

repository holds the Pf, Pf (fk..fn) as (i.e., process description, process sequence, process conditions, process performer and process data, etc.) selected by the process miner to be used in the decision-making process.

**Domain Knowledge Repository (DKR):** It contains the business regarding guidelines and knowledge to perform certain organization activities. Domain knowledge [22] is a specialized ‘background knowledge’ related to user experiences, value, and insight required to attain mastery and accuracy for a concise and clear understanding of the specific sphere’s domain. Domain experts [7] are the persons who have the expertise to solve the specific problems in their respective domain, maintained the DKR having Domain Knowledge factors (DKf), DKf (fj..fn) as (i.e., domain knowledge, domain experts, business context, etc.).

As the origin of different data causes at a different level of granularity, and the providence shows the detailed description and casual relationship among tasks and its contexts in the process [23]. The selection of relevant factors eliminates the extraneous and irrelevant factors from the dataset to improve the performance in terms of accuracy and time to build the model. The selection of appropriate factors for performing the evaluation, these repositories have been used to make smarter decisions that are extra closer to real-world business needs. Where (fi,..fn) represent DMf (i.e., data context, mining goals, technical context, etc.), (fj..fn) represents D3Mf (i.e., domain knowledge, domain experts, business context, etc.), and (fk..fn) represent Pf (i.e., process description, process sequence, process conditions, process performer and process data, etc.) at a different phase of PD3M-AKD framework.

Data miners, domain experts, process miners, and process executers work collaboratively to analyze which factors impart their part for generating erudite patterns to condense the academia-industry gap [20]. To analyze which factors impart a crucial role, calculate the support and confidence of these erudite patterns.

$$Confidence = \frac{DMF + D3MF}{D3MF}$$

$$Support = \frac{DMF}{|D|}$$

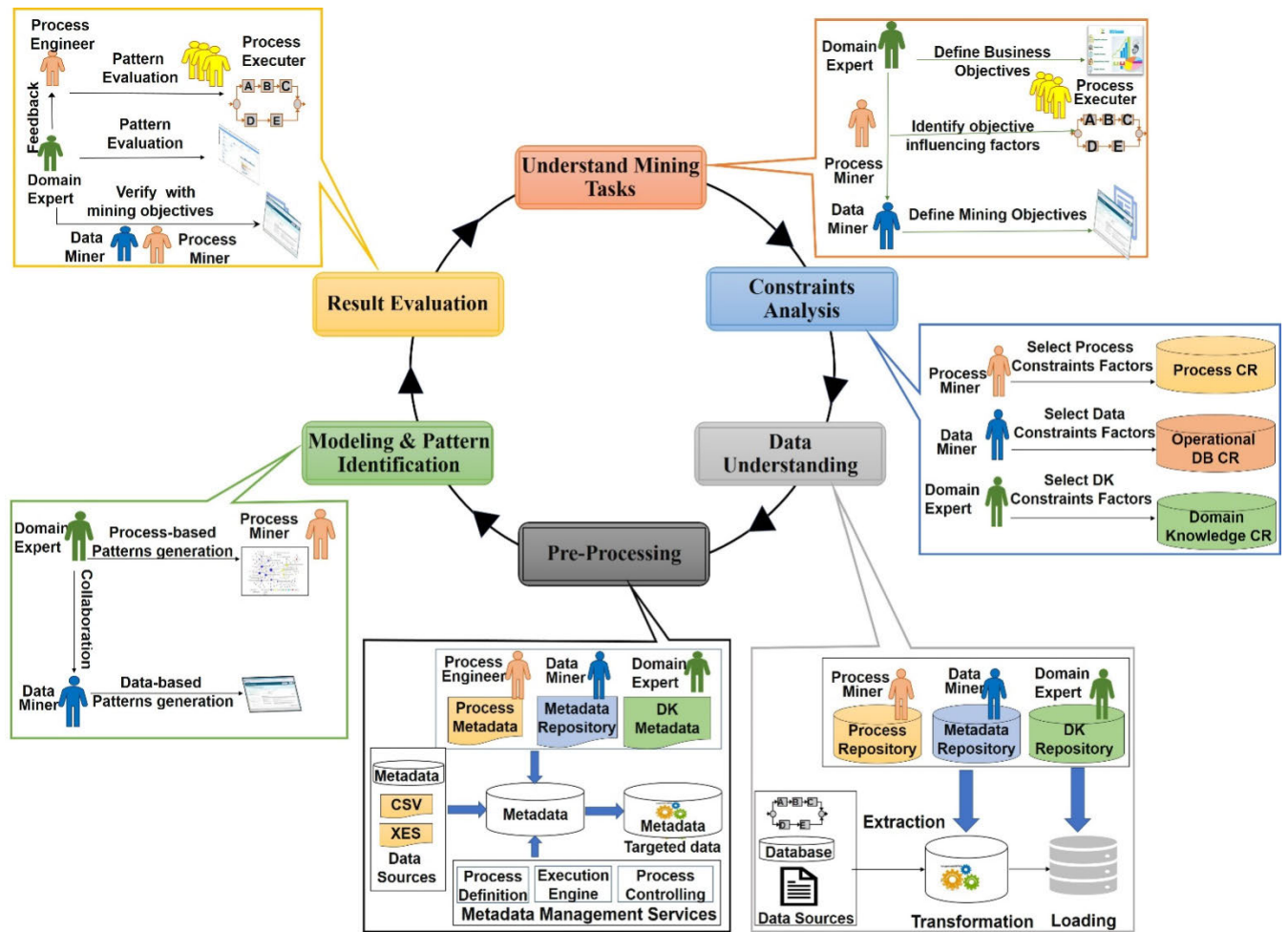


FIGURE 5. PD3M-AKD framework.

$$AKD(P) = \sum_{i,j,k} DMf_i, DKf_j, Pf_k$$

$DMf_i \in DMfN$   
 $DKf_j \in DfN$   
 $Pf \in PfN$

Where  $Pf = (Pf_{k1} + Pf_{k2} + \dots Pf_{kn})$ ,  
 $DKf = (DKf_{j1} + DKf_{j2} + \dots DKf_{jn})$ ,  
 $DMf = (DMf_{i1} + DMf_{i2} + \dots DMf_{in})$

**IV. PROPOSED PD3M-AKD FRAMEWORK**

This section describes our proposed framework, PD3M-AKD, shown in FIGURE 5. The aim of the PD3M-AKD framework is the inclusion of process-based factors from the perspectives of a business process to find more erudite applicable learned patterns. So, in this framework considerate the inclusion of data, domain, and process-based factors in the mining process that aids to overcome the inadequacy in existing phases. The availability of process log-data in the mining process opens a new era for generating erudite actionable patterns to accomplish the maximum organizational benefits. Knowledge actionability entails an essential part of the process-based factors that have not been considering in the existing D3M frameworks. Moreover, our proposed

framework facilitates eliminating the deficiencies, reviewed in section II, that causes the generation of erudite learned rules for actionable knowledge discovery.

**A. UNDERSTANDING MINING TASK**

Understanding the mining task is not straight-forward, but relatively an evolutionary, recursive, and participative process. As in the business process, various tasks have been executed in a specific order of sequence across time and place, with evidently recognized input and output to accomplish these tasks. It consists of five subprocesses:

**Determining the business objective:** Domain background knowledge, business objectives and success criteria, business scope, business interestingness.

**Assessment of resources:** Organizational resources, requirements, risks and contingencies, cost, and benefit.

**Determine data mining goals:** Data mining goals and data mining success criteria.

**Produce project plan:** Assessment of tools and techniques

**Process-based factors:** Understand the objectives by considering process-based factors, i.e., process description, process sequence, process conditions, process performer, and process data.

Accumulating these relevant factors concerning process context postulate, 4WH (What, When, Who, Where, How) while understanding the mining task for setting organizational goals. Because data, domain with the underlying process definition tell the actual story, aid in bridging the gap while understanding the mining tasks and the actual execution of the mining process to achieve the organizational goals.

### B. CONSTRAINTS ANALYSIS

For the discovery of actionable knowledge, constraints are imparting a crucial part. Our PD3M-AKD framework considered the data constraints, domain constraints, interestingness constraints, deployment constraints, and process-based constraints. Different types of constraints repositories such as the Data Constraints Repository (DCR), (i.e., data constraints, task data, block data, etc.), Domain Knowledge Constraints Repository (DKCR), (i.e., domain knowledge, domain restrictions, business rules, etc.), and the Process Constraints Repository (PCR), (i.e., process sequence, process conditions, process performer and process data, etc.) considered. Data miners, domain experts, and process miners select these constraints to achieve the business objectives.

The process-based constraints determined the actual execution order and dependencies among the tasks. For example, in [25] *sequencing*, i.e., *Task-B* can only be started when its predecessor *Task-A* has finished its execution. In *parallel split*, i.e., *Task-B* and *Task-C* can be executed concurrently after *Task-A* has finished its execution (i.e., after enrollment of fee, the student can create a student profile and course registration activities simultaneously), etc. The process is everything in the business process while executing the tasks. So, these process-based constraints help to analyze the technical, economic, and social aspects in the process while developing and deploying actionable patterns to discover actionable knowledge.

### C. DATA UNDERSTANDING

Understand the provenance of data items, according to domain knowledge and experts experience (i.e., data miner, process executor, process controller, domain experts, etc.) that include information about the process-based domain-specific data stored in the event log for analysis. Because the provenance of the data keeps track of all the internal information about the process, *data-in*, *data-out*, the *execution order of tasks*, and the task being performed, as these heavily influenced the data elements. These data elements have been integrated with the process tasks, describing which data objects have been used as input during the execution of the task (i.e., patient age, loan amount) and what the output data have been generating while task execution (i.e., outcome).

The provenance model defines the contextual information of the data stored in the event log. So, consider the process-based data patterns: i.e., data visibility, data interaction, data transfer, data-based routing patterns, etc. while understanding the data provenance. It helps to postulate

efficiency and effectiveness while mining the results because the number of process factors influences the outcomes. While performing the ETL process, DR, DKR, and PR provide the contextual provenance while understanding the data.

### D. PRE-PROCESSING ANALYSIS

At the pre-processing phase, consider the data provenance during the data extraction, transformation, and loading process as numerous process-based factors, i.e., process description, process sequence, process conditions, process performer, and process data with domain knowledge effect during the preparation of the event log. The provenance of the data stored in the process models considers the appropriate process model for the pre-processing data. The process executor engine mainly focuses on the process-parsing engine (i.e., process-definition, parsing process, and compliance checking), process-scheduler (i.e., time scheduling, tasks-dependencies), process definition that determines the metadata for maintaining the relevant factors in the event log repository.

According to the mining process's objective, data miner, domain expert, and process miner work collaboratively to select DM, D3M, and ER to analyze the most relevant factors. Where  $f$  represents the total no. of factors, and the proportion of instances that belong to class  $i$ . ( $f_i, \dots, f_n$ ) represent  $DMF$ , ( $f_j, \dots, f_n$ ) represent  $D3MF$ , ( $f_k, \dots, f_n$ ) represent different data provenance factors effect at a different phase of the PD3M-AKD framework. It helps identify the factors that impact the most significant part to determine the erudite patterns and need to be analyzed regarding domain-specific (i.e., education environment/ level), context-aware (i.e., semester system, annual system) manner for pre-processing the data.

### E. MODELING AND PATTERN IDENTIFICATION

Usually, this phase gives an analytical representation of 'as-is' processes in an organization and compares it with 'to-be' processes for making them more efficient and useful patterns. In PD3M-AKD, discover the interesting and actionable pattern using the DM algorithm and learning pattern results using manual, semi-automatic, or automating pattern recognition techniques. The rule-based pattern identification method used to provide uniform modeling patterns that emerged regarding the provenance of process, data, and domain context to examine how tasks have been executing to streamline and optimize their objectives [27].

### F. RESULT EVALUATION

PD3M-AKD results evaluation is based on considering the data, domain, and process-based factors. It determines the context that transforms 'actionable patterns' to 'erudite actionable learned rules'. This phase aims to evaluate either process-based factors impacting to achieve maximum desired outcomes or not delivering expected results while attaining actionable knowledge.

Moreover, different stakeholders, i.e., process controller, data miner, and domain experts, engaged in evaluating the results. They assist in analyzing those factors' that imparting



their role while generating the result and analyzing them regarding the business objectives [28]. Therefore, by converging these factors mentioned above, conferring with domain experts in the mining process leverages for producing more concise and actionable knowledge [29].

So, the improved accuracy of process-based patterns compared to data and domain patterns reveals that process-based factors have a dependency and played a more significant part while generating erudite actionable results.

V. CASE STUDY

Education Process Mining (EPM) combines process mining techniques with educational data to provide students, instructors, and researchers with the knowledge to benefit from the educational process. Its objective is to construct a complete and compact educational process that helps analyze students’ performance from a different context, showing different aspects of the educational environment. Three primary types of process mining techniques are discovery (deriving information from the original process), conformance (detecting, locating, and explaining deviation from the modeled process to the actual one), and extension (extended with new aspects by eradicating existing bottlenecks) [15].

In this paper, we focused on educational process monitoring and evaluation techniques using performance analysis and conformance checking by analyzing which factors impart their part to generate erudite actionable patterns to enhance business performance.

Conventional, DM techniques such as association [30], clustering [27], classification [4], regression applied to the data generated by information systems to predict the behavior and performance of the students deprived of concentrating on the underlying process being executed in the system. EDM techniques [32] are not process-centric and don’t concentrate on the event log generated data. On the other hand, EPM [33] is a process-centric approach, thereby considering the event log generated data to make the hidden patterns more erudite and actionable while taking decisions. These techniques help to analyze the students’ performance and facilitate the directors, faculty instructors, advisors, and policymakers to improve the education process [34].

However, a limited number of studies have been done in EDM and EPM to analyze which factors impact their part to improve the entire student learning process and evaluate student performance [35].

The data under the study presented, taken from the education domain of the university, Course Management System (CMS), Student Learning Management System (LMS), and Faculty Management System (FMS) of session 2014 to 2018 as shown in FIGURE 6. The CMS system represents the course selection, allocation of the instructor, enrollment of the students, class formulation and section development process, etc., and LMS includes students’ profiles data, examination results, and detailed descriptions of their learned courses. FMS includes information about faculty profile, education, specialization in a domain, courses taught and

Department Table										
Dept-Id		Dept-Name			Faculty					
Course Table										
Case Id	Course Id	Course Name	Instructor-Id	Co	Instructor Table					
Student Table				Instructor-Id	Instructor-Name	Teaching Experience	Qualification	Domain	Designation	
Std-Id	Std-Name	Ge								
1	CSE	356544301	Rabia		100	Ali Zafar	9 year	M.phil	DB	Assist.Prof
	CSE	356544302	Ayesha		101	Ahmad Aduan	28 year	PhD	SE	Professor
	CSE	356544303	Adnan		102	Sara Umar	19 year	M.phil	Programming	Assist.Prof
2	CSE	356544304	Ahmad		103	Noor ul hassan	13 year	M.phil	DB	Lecturer
	CSE	356544305	Ammara		104	Usman	18 year	PhD	SE	Professor
	CSE	356544306	Ali		105	Rabia	8 year	M.phil	Networking	Lecturer
	CSE	356544307	Ansa		106	Nasrullah Nazir	7 year	M.phil	Programming	Lecturer
3	CSE	356544308	Nida		107	Umar Hayat	5 Year	M.phil	DB	Lecturer
	CSE	356544309	Sidra		108	M.Naeem Abbas	09 years	M.phil	SE	Lecturer
	CSE	356544310	Rida		109	Sohaib Yousaf	8 year	M.phil	Networking	Assist.Prof
	CSE	356544311	Rida	101	Ms	110	Mehmood Hussan	19 year	PhD	Programming

FIGURE 6. Event log of a course management system.

expertise, etc. [36]. Each student, course, and instructor have a unique identifier (i.e., Std-id, Course-id, Instructor-Id). It generates the event log that corresponds to a set of process instances in the educational process. Each event refers to an activity that is related to a specific process instance. An event can have a timestamp (i.e., session of the semester) and a performer (i.e., an actor who performed a task).

To evaluate the usefulness of the PD3M framework, we performed an analysis of the university data set of 3 different sessions. We call them Session 1, Session 2, and Session 3. The university data must be kept anonymous. In collaboration with domain experts and process executor admin, 10 tasks regarding the courses, instructor, and students from each section, i.e., Section A, Section B, Section C, have been selected. The data set was prepared at the pre-processing stage. In section, A perform DM analysis having factors, i.e., CGPA, attendance of the student, marks. Section B for D3M analysis by integrating domain knowledge factors, i.e., subjects, their class performance, internal assessment grade, CGPA, attendance of the student, marks, etc., and in section C considered the process-based factors: process sequence, i.e., course learning order (dependency among courses), process performer, i.e., the role of instructor, who is performing the particular role, process conditions: set of courses and their relationship, how well under certain conditions by integrating domain knowledge with process-based contextual data. A typical process-flow of course allocation process, as shown in FIGURE 7.

VI. RESULTS EVALUATION

The objective is to analyze the insight knowledge of the learning process. To examine the event log data and generate erudite actionable patterns, data miner, domain experts (HOD, Advisor), and process executor (Advisor, Instructor) work collaboratively. The domain knowledge is required to determine the performance measure of a business process, assist in analyzing the event log [37].

According to our criteria, question 1, what overall actions or tasks are being performed in the overall business process. Results showed that process-based factors, i.e., what overall actions being performed in the educational process, impact a lot while evaluating students’ performance



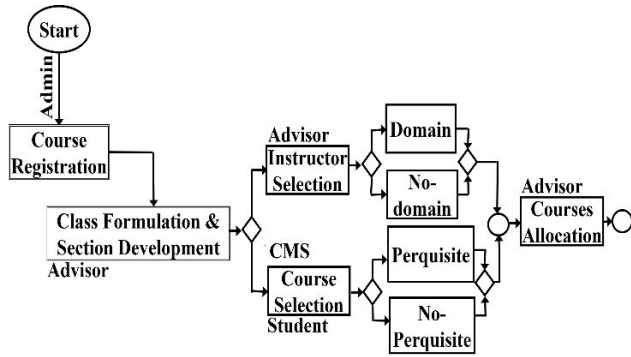


FIGURE 7. Process-flow of the course allocation process.

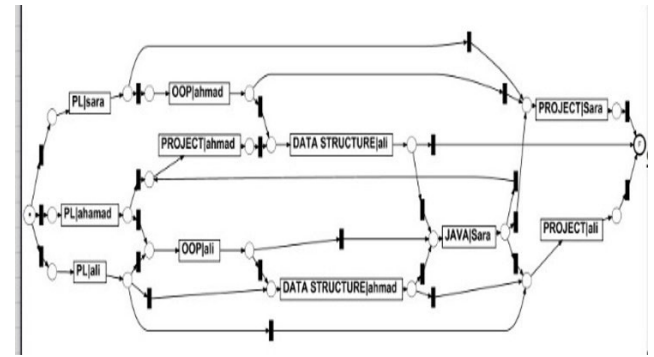


FIGURE 9. Course allocation dependency graph in Petri-net.

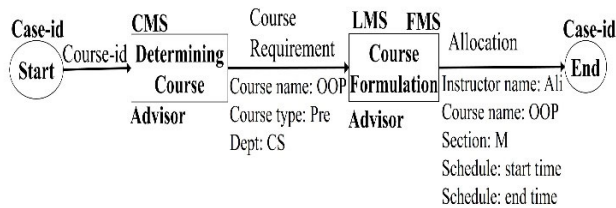


FIGURE 8. Course formulation process.

conferring to the contexts. As, in the business environment, each task is the fundamental building block of the executed process to organizational success, it played a crucial role compared to any other factor by determining the process’s contextual information [15]. According to our criteria question 2, *in which sequence different tasks are being performed*. The resulting patterns showed that the sequence of tasks and their relationship, i.e., course dependencies on other courses, have excessive influence while evaluating the students’ performance. The retain-familiar approach [32] has been used to examine the sequence in which the task has been performed, as one course depends on forthcoming courses. For example, the student’s outcome in the Programming Language (PL) course has influences on their subsequent Object-Oriented Programming (OOP) course performance. Moreover, the instructor’s role having expertise and domain knowledge in the PL course gave a better understanding to a student in clearing the concepts that impact on the student’s grades.

On the other hand, according to our criteria question 3, *by whom the tasks are being performed*, the instructor’s role played a significant role while performing a specified task. For example, having expertise in networking cannot give a good outcome in the OOP course. These results depend on the PL course’s performance taught in the previous semester and the *role* of the instructor who taught this course. Retain familiar approach helps to analyze the students’ performance in the OOP course depending on their prior task, as shown in FIGURE 8. The instructor’s expertise and their domain knowledge in the course allocation process are impacted. Because background knowledge in the previous subject and the instructor teaching method impacts the course outcome. Moreover, performing conformance checking helps to

analyze the performance of the instructor in related courses, and the instructor preferred to take the courses according to his / her expertise in their respective field.

Under certain conditions, these tasks are being performed [38]. The students’ performance is measured under the particular conditions, i.e., the execution order of the tasks, constraints followed while executing the tasks, and by whom the tasks are being performed the dependency graph of tasks and the conditions followed.

For example, in the case of class formulation, different roles (i.e., instructor) have been allocated to a case (i.e., courses) having different capabilities regarding their domain expertise [39]. There exists a dependency among courses that students have been learning in different semesters. In the next semester, when they learned advanced courses based-on these prerequisite courses [40]. Their performance depends on previous knowledge and concepts regarding the subjects they learned, directly correlated with an instructor’s expertise taught in the previous semester. Based on these *DMf*, *DKf*, and *Pf*, the following patterns have been generated as shown in (1-3) equations. Performance is evaluated by calculating the support and confidence of these selected factors.

$$[OOP = \text{“Good”} \wedge \text{Session} = \text{“M”}] \rightarrow \text{Grade} = \text{“Good”} [\text{accuracy} = 64\%] \quad (1)$$

$$[OOP = \text{“Good”} \wedge \text{Session} = \text{“M”} \wedge \text{Instructor} = \text{Ali}] \rightarrow \text{Grade} = \text{“Good”} [\text{accuracy} = 73\%] \quad (2)$$

$$[PL = \text{“Good”} \wedge \text{Instructor} = \text{“Ali”} \wedge OOP = \text{“Excellent”} \wedge \text{Session} = \text{“M”} \wedge \text{Instructor} = \text{Ali}] \rightarrow \text{Grade} = \text{“Excellent”} [\text{accuracy} = 92\%] \quad (3)$$

The resulting erudite patterns set can be processed in a semi-automated way and the process model Petri net graphs [41] to analyze the patterns of students’ ability by considering the *DMf*, *DKf*, and *Pf* as a predictor parameter as shown in FIGURE 9. To analyze the student’s behavior from the event log, the process mining ProM tool [42] is used for actual process model discovery.

The student’s performance and corresponding results help to identify the constraints that should apply during the

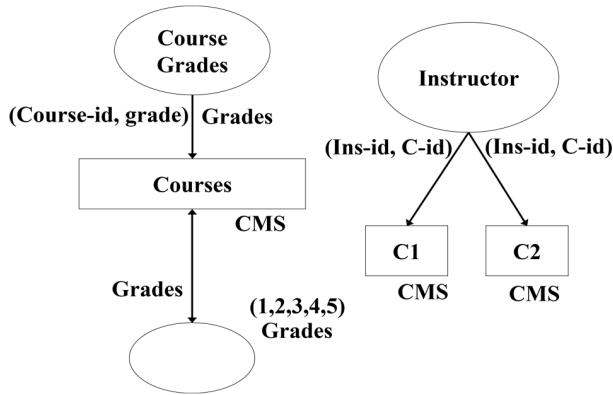


FIGURE 10. Data-in and data-out.

Wcourse allocation process. It also helps the advisors and instructors to manage their classes, analyze their teaching process, and the students’ learning process while evaluating students’ grades and performances.

According to our objective 5, *what data is provided and produced in performing these tasks*, considered the domain-specific, process generated business benefit data involvement for analysis [43]. Data elements heavily influence the execution order of tasks, that are integrated with the control-flow tasks that consider from where the input and output of the data have been generated [37].

For example, in the student evaluation process, as discussed in FIGURE 10, courses are evaluated based on the grades in the respective courses. Moreover, it defines the allocation of courses (C1 and C2) to instructors according to their expertise. FIGURE 11 discusses the allocation graph of instructors to respective courses. What dependencies of resources (i.e. instructors) are impacting the courses for generating patterns? We used a retain- familiarity approach for generating the patterns as discussed aforementioned patterns. For example, in the class formulation process, students’ results cannot be considered until they pass out in the internal evaluation, i.e., attendance, midterm, quizzes, etc. while evaluating course results. Therefore, the evaluation is based on the internal assessment as input for overall course evaluation.

To evaluate the performance, we draw a graph representing students’ performance and the instructors in the courses by patterns generated by DM algorithms. Patterns are generated in association with domain knowledge and the patterns generated by integrating domain knowledge with process-based generated data. FIGURE 12 shows the comparison results based on traditional *DMf* (GPA, semester, etc.), *DKf* (background knowledge in the domain, instructor expertise, etc.), and *Pf* is the PL course. It showed that students’ performance improved while considering process-based factors, i.e., course learning order (dependency among courses), resource allocation factors i.e. role of instructor, who is performing a certain role. during the evaluation of grades in the PL course.

FIGURE 13 shows the comparison results based on traditional *DMf* (GPA, semester, etc.), *D3M* factors (background knowledge in the domain, instructor expertise, etc.),

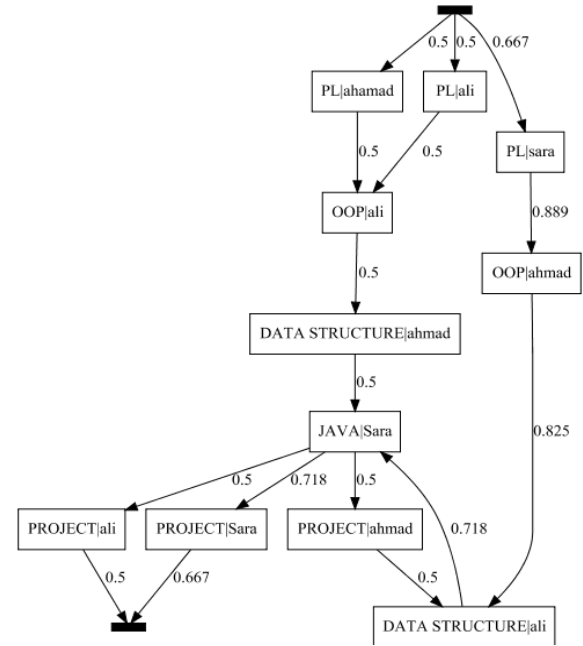


FIGURE 11. Course allocation graph.

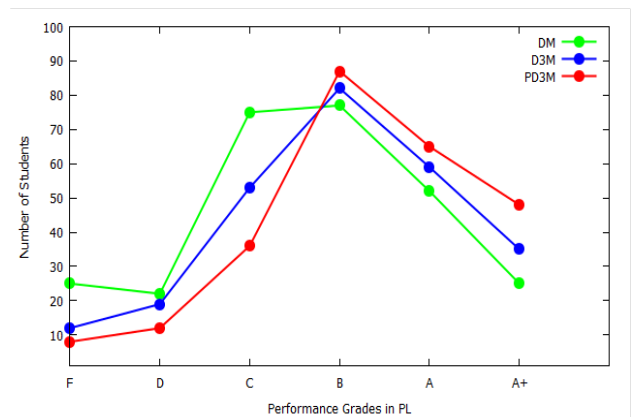


FIGURE 12. Students performance evaluation in PL.

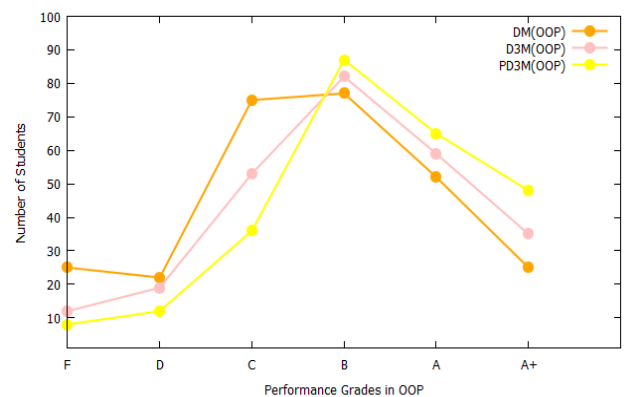


FIGURE 13. Students performance evaluation in OOP.

and *Pf* in the course of the OOP course. It showed that students’ grades performance impacted while considering *Pf*, i.e., course learning order (dependency among courses),

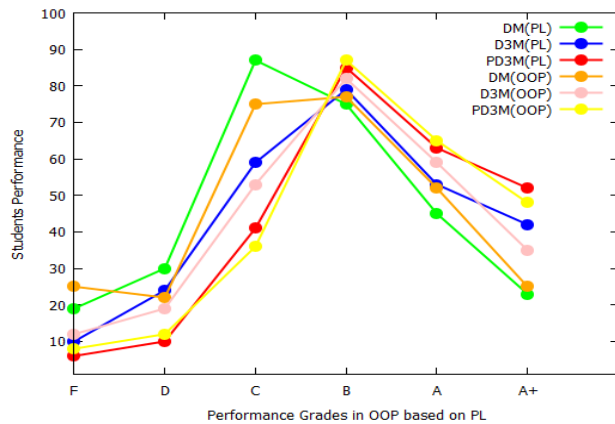


FIGURE 14. Students performance evaluation in OOP base on PL.

resource allocation, i.e., the role of instructor, who is performing a certain role (instructor) during the evaluation of grades in OOP course. FIGURE 14 compared the results of student performance in the OOP course based on retain-familiar approach on the PL course. After evaluating student performance, the instructor's role and previous knowledge of the domain have an impact than just considering  $DM_f$  and  $DK_f$ . The instructor having expertise and domain knowledge in programming subjects gave a better understanding to the student in clearing the concepts, impact the performance of the student's grades. On the other hand, the instructor having expertise in networking cannot give a good outcome in the OOP course. So, the grades depend on the PL course taught in the previous semester as well as the instructor who taught this course. Retain familiar approach help to analyze the performance of course regarding dependencies of a prior task.

## VII. CONCLUSION

In today's technological era, organizations rely on the information system to manage their business process. Our proposed PD3M framework has done a significant contribution in the data mining field by taking more appropriate actionable decisions, as it specifically deals with process relevant information according to business objectives. The contribution of this paper is multi-fold. Firstly, the decisions taken by entailing these process-based factors according to data and domain knowledge, make erudite patterns more actionable. As the process is nothing without data, and data is meaningless without the process and its context. So, ignoring these process-based factors, the generated patterns may be innovative and efficient but not actionable to be applicable in the real-world. Secondly, there is a clear recognition for the domain experts and business experts in the phases of the PD3M framework for erudite actionable patterns generation. as they faced the consequences of decisions taken during the mining process.

Therefore, incorporating these process-based domain knowledges concerning factors with the consideration of domain experts in mining process phases, makes learned patterns more actionable and interesting for enchanting smarter decisions compatible with real-world business needs. A case study from the educational domain has been presented to

evaluate our results, showing how process-based factors are used along with data and domain knowledge factors to model, analyze, and improve the mining process. After evaluation, results show that learned rules actionability improved when process relevant factors were considered from the above five perspectives of a business process compared to the rules learned merely considering factors only from dataset or domain knowledge. Our proposed framework helps to eliminate the deficiencies, causing the generation of erudite patterns for actionable knowledge discovery. Future research could evaluate how process-based patterns yield a better understanding of the implications of taking actionable decisions.

## REFERENCES

- [1] L. Cao and C. Zhang, "Domain-driven data mining: A practical methodology," *Int. J. Data Warehousing Mining (IJDDWM)*, vol. 2, no. 4, pp. 49–65, 2006.
- [2] L. Cao, D. Luo, and C. Zhang, "Knowledge actionability: Satisfying technical and business interestingness," *Int. J. Bus. Intell. Data Mining*, vol. 2, no. 4, pp. 496–514, 2007.
- [3] L. Cao and C. Zhang, "The Evolution of KDD: Towards domain-driven data mining," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, no. 4, pp. 677–692, Jun. 2007.
- [4] X. Zhao and S. Mafuz, "Towards incorporating context awareness into business process management," *Int. J. social, Behav., Educ., Econ., Bus. Ind. Eng.*, vol. 9, no. 12, pp. 3890–3897, 2015.
- [5] M. Wieland, O. Kopp, D. Nicklas, and F. Leymann, "Towards context-aware workflows," in *Proc. CAISE Workshops Doctoral Consortium*, Jul. 2007, vol. 2, no. S25, p. 15.
- [6] L. Cao and C. Zhang, "Domain-driven actionable knowledge discovery in the real world," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, Apr. 2006, pp. 821–830.
- [7] Y. Shi, L. Zhang, Y. Tian, and X. Li, *Data Mining Integrated With Domain Knowledge Intelligent Knowledge: A Study Beyond Data Mining*. Berlin, Germany: Springer, 2015.
- [8] O. Alejandro Molina and S. Ratté, "Students and teacher academic evaluation perceptions: Methodology to construct a representation based on actionable knowledge discovery framework," *Edu. Inf. Technol.*, vol. 22, no. 3, pp. 1043–1066, May 2017.
- [9] A. Huang, L. Zhang, Z. Zhu, and Y. Shi, "Data mining integrated with domain knowledge," in *Cutting-Edge Research Topics on Multiple Criteria Decision Making*. Berlin, Germany: Springer, 2009, pp. 184–187.
- [10] F.A. Rahman, S.M. Shamsuddin, S. Hassan, and, N. A. Haris, "A review of KDD-data mining framework and its application in logistics and transportation," *Int. J. Supply Chain Manage.*, vol. 5, no. 2, pp. 77–84, 2016.
- [11] U. Fayyad, S. G. Piatetsky, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 3, no. 17, p. 37, 1996.
- [12] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019.
- [13] A. I. R. L. Azevedo and M. F. Santos, "KDD, SEMMA and CRISP-DM: A parallel overview," in *Proc. IADS-DM*, 2008, pp. 182–185.
- [14] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *Knowl. Eng. Rev.*, vol. 25, no. 2, pp. 137–166, Jun. 2010.
- [15] F. Martinez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernandez Orallo, M. Kull, N. Lachiche, M. J. Ramirez Quintana, and P. A. Flach, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Trans. Knowl. Data Eng.*, early access, Dec. 27, 2019, doi: 10.1109/TKDE.2019.2962680.
- [16] T. Logan, "A practical, iterative framework for secondary data analysis in educational research," *Austral. Educ. Researcher*, vol. 47, no. 1, pp. 129–148, Mar. 2020.
- [17] H. J. Palacios, R. A. J. Toledo, G. A. H. Pantoja, and A. Navarro, "A comparison between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 2, no. 3, pp. 598–604, 2017.



- [18] G. Milosavljevic, G. Sladic, B. Milosavljevic, M. Zaric, S. Gostojic, and J. Slivka, "Context-sensitive constraints for access control of business processes," *Comput. Sci. Inf. Syst.*, vol. 15, no. 1, pp. 1–30, 2018.
- [19] L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E. K. Park, "Flexible frameworks for actionable knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 9, pp. 1299–1312, Sep. 2010.
- [20] F. Piazza and S. Strohmeier, "Domain-driven data mining in human resource management: A review," in *Proc. IEEE 11th Int. Conf. Data Mining Workshops*, Dec. 2011, pp. 458–465.
- [21] C. Schmidt, C.-P. Praeg, and J. Gunther, "Designing digital workplace environments," in *Proc. IEEE Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*, Jun. 2018, pp. 1–9.
- [22] H. A. Reijers, M. Song, and B. Jeong, "Analysis of a collaborative workflow process with distributed actors," *Inf. Syst. Frontiers*, vol. 11, no. 3, pp. 307–322, Jul. 2009.
- [23] J. Cardoso, R. P. Bostrom, and A. Sheth, "Workflow management systems and ERP systems: Differences, commonalities, and applications," *Inf. Technol. Manage.*, vol. 5, no. 3/4, pp. 319–338, Jul. 2004.
- [24] M. Wang and H. Wang, "From process logic to business logic—A cognitive approach to business process management," *Inf. Manage.*, vol. 2, no. 43, pp. 179–193, Mar. 2006.
- [25] T. T. P. Thi, M. Helfert, F. Hossain, and T. L. Dinh, "Discovering business rules from business process models," in *Proc. 12th Int. Conf. Comput. Syst. Technol. (CompSysTech)*, 2011, pp. 259–265.
- [26] L. Grumbach and R. Bergmann, "Workflow flexibility by deviation by means of constraint satisfaction problem solving," *Proc. LWDA*, Dec. 2017, p. 212.
- [27] M. Pesic, M. H. Schonenberg, N. Sidorova, and W. M. Aalst, "Constraint-based workflow models: Change made easy," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Berlin, Germany: Springer, Nov. 2007, pp. 77–94.
- [28] M. Zerari and M. Boufaïda, "Dynamic context-aware Business Process flexibility: An artefact-based approach using process mining," *Int. J. Bus. Intell. Data Mining*, vol. 4, no. 6, pp. 345–361, 2011.
- [29] M. L. Eck, X. Lu, S. J. Leemans, and W. M. Aalst, "PM<sup>2</sup>: A process mining project methodology," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Cham, Switzerland: Springer, Jun. 2015, pp. 297–313.
- [30] R. Talib, B. Volz, and S. Jablonski, "Agent assignment for process management: Goal modeling for continuous resource management," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, Sep. 2010, pp. 25–36.
- [31] N. Russell, W. M. Aalst, A. H. Hofstede, and D. Edmond, "Workflow resource patterns: Identification, representation and tool support," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Berlin, Germany: Springer, Jun. 2005, pp. 216–232.
- [32] W. M. P. Aalst, "Business process management," *Bus. Inf. Syst. Eng.*, vol. 1, no. 58, pp. 1–6, 2016.
- [33] R. Talib, B. Volz, and S. Jablonski, "Agent assignment for process management: Agent performance evaluation framework," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2010, pp. 1005–1012.
- [34] A. Meyer, S. Smirnov, and M. Weske, "Data in business processes," Universitätsverlag, Univ. Potsdam, Potsdam, Germany, Tech. Rep. 50, 2011.
- [35] W. M. P. Aalst, "Business process management: A comprehensive survey," *ISRN Softw. Eng.*, vol. 2013, p. 37, May 2013.
- [36] R. Batra and M. A. Rehman, "Actionable knowledge discovery for increasing enterprise profit, using domain driven-data mining," *IEEE Access*, vol. 7, pp. 182924–182936, 2019.
- [37] L. Cao and C. Zhang, "Domain Driven Data Mining," in *Data Mining and Knowledge Discovery Technologies*, vol. 9, no. 2 Hershey, PA, USA: IGI Global, 2008, pp. 196–223.
- [38] C. A. M. Soares, Y. Peng, and J. Meng, *Applications of Data Mining in E-business and Finance*, vol. 177. Amsterdam, The Netherlands: Ios Press, 2008.
- [39] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1355, 2020.
- [40] S. Berti, J. Zelst, and W. V. Aalst, "Process mining for python (PM4PY): Bridging the gap between process-and data science," 2019, *arXiv:1905.06169*. [Online]. Available: <https://arxiv.org/abs/1905.06169>
- [41] J. Li, S. H. Han, and S. Fu, "Exploring the relationship between students' learning styles and learning outcome in engineering laboratory education," *J. Further Higher Educ.*, vol. 8, no. 43, pp. 1064–1078, 2019.
- [42] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.
- [43] C. Chang, L. He, N. Chaudhary, S. Fu, H. Chen, J. Sun, K. Li, Z. Fu, and M.-L. Xu, "Performance analysis and optimization for workflow authorization," *Future Gener. Comput. Syst.*, vol. 67, pp. 194–205, Feb. 2017.



**FAKEEHA FATIMA** is currently pursuing the Ph.D. degree with the Department of Computer Science, Government College University, Faisalabad. Her research interests are databases, data mining, and process mining.



**RAMZAN TALIB** received the Ph.D. degree from the University of Bayreuth, Germany. He is currently working as a Professor with the Department of Computer Science, Government College University, Faisalabad. His research interests include databases and information systems, data mining, data warehousing, business process management, and workflow management systems.



**MUHAMMAD KASHIF HANIF** received the Ph.D. degree from the Hamburg University of Technology. He is currently working as an Assistant Professor with the Department of Computer Science, Government College University, Faisalabad. His research interests include big data analytic, scientific computing, and bioinformatics.



**MUHAMMAD AWAIS** received the Ph.D. degree from the University of Bayreuth, Germany. He is currently working as an Assistant Professor with the Department of Software Engineering, Government College University, Faisalabad, Pakistan. His research interests are automation and robotics and artificial intelligence.