Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

A Rough Set–Based Process Mining Approach for Knowledge Intensive Processes: Insights from a New Product Development Case Study

Mehri Chehrehpak*

*Corresponding author, Ph.D. Candidate, Faculty of Management and Economy, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: mehri.chehrehpak@gmail.com

Abbas Toloui Ashlaghi

Professor, Faculty of Management and Economy, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: toloei@srbiau.ac.ir

Kamran Mohammadkhani

Professor, Faculty of Management and Economy, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: k.kamran@srbiau.ac.ir

Journal of Information Technology Management, 2026, Vol. 18, Issue 3, pp. 25-46
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.402437.4270
Article Type: Research Paper

Received: January 13, 2026
Received in revised form: March 09, 2026
Accepted: May 21, 2026
Published online: July 01, 2026

Abstract

Knowledge-intensive processes (KiPs) have gained significant attention in modern organizations, where decision-making plays a fundamental role in the value chain. Therefore, the analysis of decision-making logs and the identification of decision rules and models are crucial. This paper presents a decision model for optimizing the New Product Development (NPD) process based on a case study at a prominent information technology company in Iran. The study leverages rough set theory and a fast reduction algorithm to analyze decision points within the NPD process. The model introduces a step-by-step methodology for identifying decision models at critical decision points, aiming to streamline decision-making and enhance process efficiency. The paper discusses the identification of critical decision features and the creation of decision trees for each decision point, thereby illuminating the impact of the decision model. Moreover, the study reports a reduction in the number of decisions and improved process efficiency, with a substantial 90% reduction recorded after implementing the model in a real-world scenario within the organization. The paper also highlights certain limitations and proposes future research directions, emphasizing the potential for applying the model to other knowledge-intensive processes and within larger organizational contexts. The

comprehensive review of the case study and the proposed decision model offer valuable insights into enhancing decision-making processes within the new product development domain.

Keywords: Process Mining, Decision Mining, Rough Set Theory, knowledge-intensive process, New Product Development, Information Technology

Introduction

Business Process Management (BPM) is a scientific field that has captured the attention of scholars for many years, offering numerous applications in real-world organizational management (Weske, 2019). In recent years, there have been significant advancements in BPM approaches and tools, with various methods proposed for process modelling (Weber et al., 2009). Traditional process management approaches have been primarily developed on the basis of the assumption that processes consist of recurring tasks with well-defined execution flows (Marjanovic & Freeze, 2011). Consequently, these models are most suitable for production and administrative processes (Leymann & Roller, 1999). However, many real-world processes, such as healthcare procedures, emergency management, project coordination, and research and development (R&D) processes in production units, do not fit this classical model. These processes, which are largely unstructured or less structured, fall under the category of knowledge-intensive processes (KiP) (Reichert & Weber, 2012). In today's competitive environment, where knowledge creation, retention, and distribution within organizations are of paramount importance, scholars have increasingly focused on knowledge-intensive processes (Weske, 2019; Mazorodze, 2020; Yin et al., 2022). Moreover, a significant majority of value-creating processes in the information technology field, such as the development of innovative IT-based services and new product production, are also categorized as knowledge-intensive processes (Bolisani & Scarso, 1999; Nikabadi & Sepehrnia, 2019).

There has been a significant shift in organizations regarding the use of process data in recent decades. This shift, commonly known as "Big Data Analytics (BDA), " is driving organizations toward real-time integration of business data and its analysis to gain organizational value. One of the most important BDA techniques that can bring significant benefits to organizations is process mining, which aims to discover, monitor, and improve business processes by analyzing event data generated and stored in information systems (Van der Aalst, 2011). The field of process mining is constantly growing in terms of available research and can bring significant benefits to organizations (Heudecker & Hare, 2016). However, it should be noted that the analysis of process data and logs is not a simple task (Pedro, Brown, & Hart, 2019; Baesens et al., 2016), and according to reports, a significant percentage of organizations are unable to complete these tasks (Heudecker & Hare, 2016). Therefore, having large amounts of data in the current world does not create value. Instead,

conducting a proper analysis of the data leads to strategic value. Hence, numerous studies have been conducted on the effectiveness of BDA and process mining (Grover et al., 2018; Constantiou & Kallinikos, 2015).

In recent years, research in the field of process mining has received special attention because of the increasing importance of knowledge-intensive processes and the knowledge derived from these processes. It is expected that the volume of research in this area will continue to grow in the future (Reder et al., 2019). For example, Van Dongen, De Medeiros, and Wen (2009), Wang et al. (2012), De Roock and Martin (2022), and Van der Aalst (2022) are noted. Some scholars have also focused on the applications of process mining in various fields, such as healthcare (e.g., De Roock & Martin, 2022; Ghasemi & Amyot, 2016), elderly care (Farid, De Kamps, & Johnson, 2019), e-learning (Ghazal, Ibrahim, & Salama, 2017), supply chain management (Jokonowo et al., 2018), and information technology (Fauzi & Andreswari, 2022; Urrea-Contreras et al., 2021). Norozi et al., (2026) shows that "Insecure Firmware/Software and Inadequate Patch Management" is the top cybersecurity risk, followed by "Lack of Standardization and Interoperability Issues" and "Physical Security concerns". Also Ebrahimi Kordlar and Zakeri (2009) investigated earnings management through the sale of firm assets. One of the factors that previous studies have identified as an influential factor in Earning Response Coefficient (ERC) is default risk (Ebrahimi Kordlar & Mohammadi Shad, 2014). Ebrahimi Kordlar and Aarabi (2010) review the relationship between ownership concentration and earnings quality by testing the data collected from 148 listed companies in the Tehran's Stocks Exchange.

According to Rozinat and van der Aalst (2006), most studies in the field of process mining and developed algorithms focus on process discovery and its compatibility from the perspective of control flow, while paying less attention to data features that are effective in process execution routing. However, recent advances have resulted in the development of methods for evaluating decisions during processes (De Leoni & Van der Aalst, 2013). These methods, known as decision mining, enable the discovery of decision points to explain how different routes are taken during process execution. Decision points are "parts of a process model where the process is split into alternative branches" (Rozinat & Van der Aalst, 2006).

It should be noted that for decision mining and identifying decision states at each decision point, more detailed information about the decision-making context for each case is needed in addition to process logs. For this purpose, the concept of a decision log is commonly used. The decision log can be obtained from information sources such as business management systems, ERP systems, and similar resources (Zhu et al., 2015). Different data mining algorithms and techniques, such as decision trees, random forests, k-nearest neighbor, support vector machines, Bayesian classification, and artificial neural networks, have been used by researchers to analyze decision logs (Gupta et al., 2017; Aggarwal & Zhai, 2012). Decision mining has been utilized by researchers in various domains, such as healthcare (Munoz-Gama

et al., 2022; Moreira et al., 2019; Mishra, 2019), software development processes (Portolani et al., 2023), banking systems (Srivastava, 2021), insurance process analysis (Poppe et al., 2021), and construction processes (Mohemad et al., 2010).

This study proposes a model for analyzing decision logs and modeling decision-making in knowledge-intensive processes. To achieve this aim, the focus will be on the new product development process in the field of information technology within one of the largest Iranian IT companies. The available data in the company databases will be analyzed. Initially, the critical decision features with the highest impact on decision outputs will be identified. Then, a rule-based decision model is constructed using if-then rules. To this end, the rough sets theory, which is one of the commonly used methods in data mining (Pawlak et al., 2001), and the fast reduction algorithm developed on the basis of this theory will be used. A step-by-step method for decision modeling using the decision log is presented in the form of a decision tree.

The rest of the paper is organized as follows. Section 2 reviews the case study and the process of new product development in this case. Section 3 discusses the methodology of decision point identification, the decision rule identification algorithm, and rough set theory. Section 4 will explain the proposed model step by step. Finally, Section 5 presents conclusions and recommendations for future research.

Literature Review

Behpardaz Jahan, with more than 25 years of experience in the field of information technology, is one of the leading companies in this field in Iran. During these years, this company has successfully completed over 500 projects in various areas, such as e-government, deployment of business support systems in mobile operators, e-health, and launching new services in the B2B and B2C sectors. This company now has more than 800 employees. Since 2008, the company has been using various process management tools to shape its processes. The workflow within the organization is mechanized, and as a result, a large volume of process logs exists within the company.

One of the processes that the company has been seriously pursuing since 2011 is new product development studies. For this purpose, a business development unit has been established within the company, which currently employs nearly 40 specialized personnel in this field. The primary mission of this team is to conduct feasibility studies for the ideas proposed in this field, which may come from colleagues within the company or from individuals outside the organization. Figure 1 presents the process of new product development in this company.

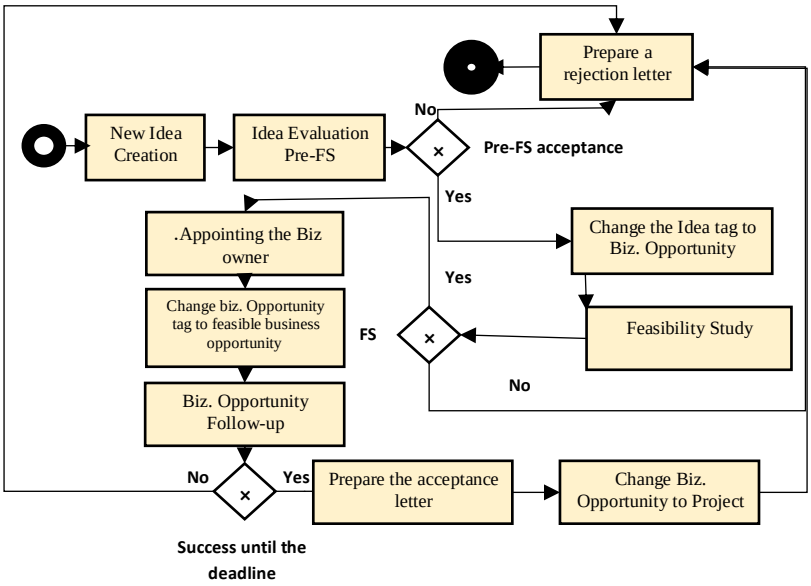

Figure 1. Process of project or business model formation in the company

As Figure 1 shows, an idea is initially proposed for which a profile is formed by the business development team in the form of Table 1.

Table 1. Summarized idea profile

No.	Title	Acceptable value	Symbol[1]
1	Idea source	1) employees; 2) board members; 3) consultants; 4) others.	V_1
2	Idea Presenter Creditability	1) high; 2) moderate; 3) low.	V_2
3	Idea relevance to existing businesses	1) high; 2) moderate; 3) low.	V_3
4	Idea relevance with existing customers	1) high; 2) moderate; 3) low.	V_4
5	Idea innovativeness	1) high; 2) moderate; 3) low.	V_5
6	Idea relevance with available technologies	1) high; 2) moderate; 3) low.	V_6
7	Idea implementation	1) EPC project; 2) BOT implementation; 3) BLT implementation; 4) e-service development.	V_7

In the next step, pre-feasibility studies will be conducted on the presented idea. Because of these pre-feasibility studies, which are conducted by the business development team, market research is carried out, and six new information fields are added to the idea profile, which are listed in Table 2.

[1] The presented symbol for decision modeling will be used later.

Table 2. Summary of pre-feasibility studies

No.	Title	Acceptable value	Symbol
1	Idea novelty	1) brand new; 2) Experience in other countries; 3) Experience in the country.	V_8
2	Number of competitors	1) high; 2) limited; 3) none.	V_9
3	Expected customer acceptance	1) high; 2) moderate; 3) low.	V_{10}
4	Need for a new license	1) No need;2) Can start without a license and obtain it afterward; 3) Requires a license that can be easily obtained; 4) Obtaining a license is very difficult.	V_{11}
5	Development risk	1) high; 2) moderate; 3) low.	V_{12}
6	Implementation risk	1) high; 2) moderate; 3) low.	V_{13}

The studies conducted are now presented to the portfolio committee for decision-making on whether to continue the studies or archive the idea. The portfolio committee is the highest decision-making authority within the company, consisting of members of the Board of Directors and some of their advisors. If the idea is not approved, it is documented and the idea presenter is notified. However, if the idea is approved, it is transformed into a business opportunity, and the feasibility study phase begins. In feasibility studies, for each business opportunity, the parameters presented in the initial studies are further elaborated upon, and based on various parameters, the revenue and financial aspects of the project are presented. Table 3 provides a summary of the feasibility studies for each idea.

Table 3. Summary of feasibility studies

No.	Title	Acceptable value	Symbol
1	IRR calculated	1) Less than 10% 2) 10%–20% 3) 20%–30% 4) More than 30%	V_{14}
2	NPV with 5% interest rate	1) Less than $100,000 2) $100,000 to $300,000 3) $300,000 to $500,000 4) More than $500,000	V_{15}
3	Payback period	1) More than 5 years 2) 3 to 5 years 3) 1 to 3 years 4) Less than one year	V_{16}
4	Net profit	1) Less than 20% 2) 20%–30% 3) 30% to 50% 4) More than 50%	V_{17}
5	Production time	1) More than 3 years 2) 2 to 3 years 3) 1 to 2 years 4) Less than 1 year	V_{18}
6	Capital Expenditure (CAPEX)	1) Less than $300,000 2) $300,000 to $500,000 3) $500,000 to $1,000,000 4) More than $1,000,000	V_{19}
7	Operating Expenses	1) Less than $300,000 2) $300,000 to $500,000 3) $500,000 to $1,000,000 4) More than $1,000,000	V_{20}
8	Maximum required capital	1) Less than $300,000 2) $300,000 to $500,000 3) $500,000 to $1,000,000	V_{21}

No.	Title	Acceptable value	Symbol
		4) More than $1,000,000	
9	Number of specialized workforces required for the production cycle	1) Less than 10 people 2) 10 to 30 people 3) More than 30 people	V_{22}
10	Number of unspecialized workforces required for the production cycle	1) Less than 10 people 2) 10 to 30 people 3) More than 30 people	V_{23}
11	Number of specialized workforces required for the implementation phase	1) Less than 10 people 2) 10 to 30 people 3) More than 30 people	V_{24}
12	Number of unspecialized workforces required for the implementation phase	1) Less than 10 people 2) 10 to 30 people 3) More than 30 people	V_{25}
13	Hardware cost	1) No hardware required 2) Less than 10% of the project cost 3) 10%–30% of the project cost 4) More than 30% of the project cost	V_{26}

After the feasibility study phase, the portfolio committee reconvenes and reviews the results. The review result can either be the closure of the business opportunity or its acceptance. If the opportunity is not approved, the archived opportunity is communicated back to the idea presenter. If the opportunity is approved, an owner will be assigned to the opportunity, and the opportunity owner, along with the business unit and other departments, will proceed with converting the opportunity into a project. During this phase, activities such as data verification, prototyping, business negotiations, and acquiring necessary licenses are conducted. If the set goals are not achieved within the timeframe determined by the committee for converting the opportunity into a project, the committee may proceed with archiving the opportunity. Otherwise, after taking necessary actions and gaining the approval of the portfolio committee, the business opportunity is converted into a project, and the process ends.

Based on the abovementioned process since the beginning of 2011, 923 ideas have been submitted to the business secretariat, out of which 409 ideas have been converted into business opportunities. Out of the identified business opportunities, the portfolio committee has accepted 156 opportunities, of which 38 new projects have been created. The final point is that, given the time frame of the process and considering changes in the country's macroeconomic parameters for evaluating business opportunities, the study period has been added as a tag in the relevant opportunity database. These study periods were classified as follows: 1) 2011 to 2015 (the period of sanctions before JCPOA), 2) 2015 to 2018 (the period of accepting JCPOA until the US withdrawal from this agreement), and 3) from 2018 until now (the period after the US withdrawal from JCPOA).

Methodology

Framework

This study focuses on the identification of decision points in processes based on the framework proposed by Bazhenova and Weske (2016). Accordingly, the key concepts are defined.

Definition 1 (Process Model): Process model $m = (N, D, \Sigma, C, F, \alpha, \xi)$consists of a finite non-empty set N of control flow nodes, a finite set D of data nodes, a finite set Σ of conditions, a finite set C of directed control flow edges, and a finite set F of directed data flow edges. The set of control flow nodes $N = J \cup E \cup G$consists of mutually disjoint $J \subseteq T \cup S$ of activities being tasks T or subprocesses S, set E of events, and set G of gateways. $C \subseteq N \times N$ is the control flow relation such that each edge connects two control flow nodes. $F \subseteq (D \times J) \cup (J \times D)$ is the data flow relation indicating read and write operations of an activity with respect to a data node. Let Z be a set of control flow constructs. Function $\alpha: G \to Z$ assigns to each gateway a type in terms of a control flow construct. Function $\xi: (G \times N) \cup C \nrightarrow \Sigma$ assigns conditions to control flow edges originating from gateways with multiple outgoing edges.

Definition 2 (Decision point): Let $m = (N, D, \Sigma, C, F, \alpha, \xi)$ be a process model. Then, $d_p = (N', D', \Sigma', C', F', \gamma, \sigma)$ is a decision point for the process model m if

1. d_p is a connected subgraph of the process model, m, where $N' \subseteq N$, $D' \subseteq D$, $\Sigma' \subseteq \Sigma$, $C' \subseteq C$ and $F' \subseteq F$;
2. The functions γ and σ are restricted functions of α and ξ with new domains.
3. $|G_{dp}| = 1 \wedge |g\bullet| \geq 2 \wedge (\gamma(g) = XOR \vee \gamma(g) = IOR)$ and $g \in G_{dp}$ (The decision point consists exactly of a split gateway);
4. $|J_{dp}| = |g\bullet| + 1$ (The number of activities of $d_p$equals the number of outgoing edges of the split gateway g plus 1[2]);
5. $\bullet g = t \wedge |\bullet g| = 1, t \in T_{dp}$ (Task t is the only predecessor of the split gateway)
6. $|\bullet t| = 0$ (Task t, is the initial node d_p).
7. $\forall j \in J_{dp} \backslash t: j \bullet = g$ (All activities other than the one preceding the split gateway g directly succeed g)
8. $\forall j \in J_{dp} \backslash t: |j \bullet| = 0$ (All activities other than the one preceding the split gateway g directly succeed g)

[2] one input task plus the number of output tasks

9. $\forall j \in J_{dp} \backslash t, c \in C_{dp}$ where $(g, a) = c: \sigma(c) \in \Sigma_{dp}$ (all outgoing edges of the split gateway are annotated with a condition)

Figure 2 presents the three decision points related to the process shown in Figure 2. In all three decision points, the decisions are of the XOR type; at the first decision point (dp1), the options are acceptance and changing the idea to a business opportunity, and in case of rejection, sending a letter to reject the decision outcomes. The same is the case with dp2 and dp3. Moreover, only two output control routes exist in all three decision points ($|g \bullet| = 2$).

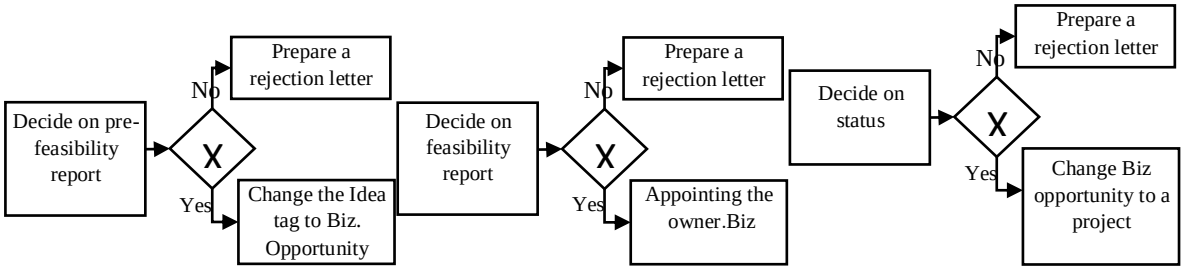

Figure 2. Triple decision points for the process presented in Figure 2

Since decisions are made within the scope of process models, it is important to consider the context related to the process. To infer decisions, according to Bazhenova & Weske (2016), a specific decision rule framework is used to consider measurable criteria called key performance indicators (KPIs). This approach assumes that the values of KPIs are associated with single tasks of all decision points in a process model.

Definition 3 (KPIs of Alternatives in a Decision Point): Let $DP = \{d_{p1} \dots d_{pM}\}$ be a set of M decision points of a process model m. Then $K = \{K_1 \dots K_M\}$ is a set of $i \in \mathbb{N}^+$ key performance indicators (KPIs) if there is a function $\psi: (t_{dpj} \bullet \times K) \nrightarrow \mathbb{R}, 1 \leq j \leq M$, which assigns the value of a given KPI to the tasks succeeding a gateway (alternatives) in each of the decision points of the process model. Bazhenova and Weske (2016) proposed enriching the business rules obtained by decision mining by considering the values of the KPIs corresponding to the activities of a decision point.

Definition 4 (Decision Table): The decision table dt = (I, O, R) consists of a finite nonempty set I of inputs, a finite non-empty set O of inputs, and a list of rules R, where each rule is composed of the specific input and output entries of the table row.

According to these definitions, Bazhenova and Weske (2016) presented an algorithm for inferring a decision model and adapting it to a process model (Figure 3).

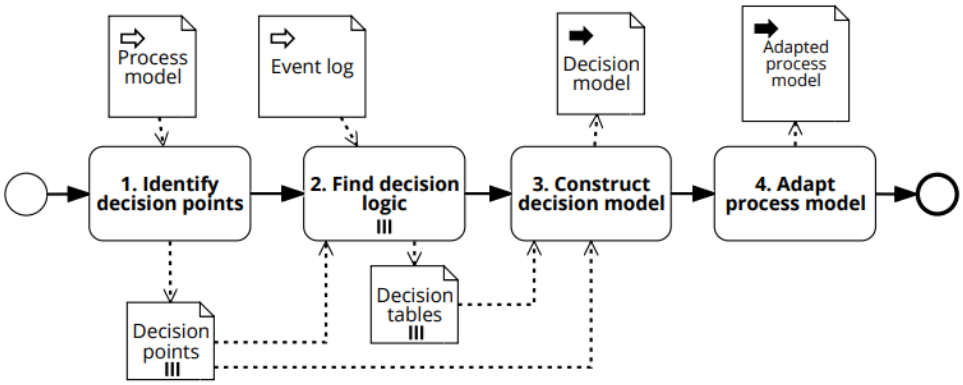

Figure 3. Algorithm for inferring a decision model and adapting it to a process model (Bazhenova & Weske, 2016)

As shown in Figure 3, the algorithm identifies and adapts the decision model to the process model in four steps:

Step 1- Determining decision points (dpi)

Step 2- Identifying decision logic: based on the decision table for each decision point, attempts are made to identify effective factors in the decision-making process based on the collected decision logs. In this study, we use the rough set theory and fast reduction algorithm.

Step 3- Constructing the decision model: based on the factors identified in the previous step and by constructing a decision tree, we present this model.

Step 4- Adapting the process model: After extracting the decision logic from the process model to the decision model, the process model must be adapted to the decision model. For this purpose, BPMN provides business rule tasks that can be linked to relevant decision models to obtain the high-level decision value of the decision model (Hosic et al., 2020).

Rough Set Theory

The problem of identifying critical decision-making features among n existing features in the decision log and forming a decision tree can be modelled as a dimensionality reduction problem, and data dimension reduction methods can be used to solve it. Data dimension reduction is an important method in the process of knowledge discovery, especially when dealing with a large amount of data. The theory of rough sets is one of the methods used in the dimensionality reduction process (Pawlak, 1991). For each decision point, the matrix $PIE_{m\times(n+1)} = [C_{m\times n} \quad D_m]$ represents the status of n decision-making features recorded for m different options (business ideas and opportunities) and decision outcomes (such as acceptance or rejection). The parameters that have the most significant effect on the decision outcome are referred to as critical decision features. Critical decision features are a subset of

decision-making features. By controlling these features, the decision outcome can be predicted for each idea or business opportunity. The theory of rough sets is applied to the PIE table. Each column of the matrix represents the status of a feature for different options, and each row summarizes the status of each option based on available features. Table 5 presents a section of the matrix for the first decision point.

The PIE matrix is defined as a quadruple $S =< U, A = C \cup D, \{V_a | a \in A\}, \{f_a | a \in A\})$, where U is a non-empty set of records, A is a non-empty set of features and outcomes. Set A can be defined as the union of features (C) and outcomes (D). V_a is a non-empty set of values for each feature, and $f_a: U \rightarrow 2^{v_a}$ is an information function for feature $a \in A$. Using rough sets theory, a subset of features can be found that play a fundamental role in producing an outcome (in this case: critical decision features). The basis for finding these features is the concepts of indiscernibility, lower set approximation, and reduct, which are defined in the theory of rough sets (Ziarko, 1993).

Definition 5 (indiscernibility): Two records, x and y, are considered indiscernible with respect to feature a (denoted as $xR_a y$) if and only if they have the same values for this feature. In other words, $\forall x, y \in U,\ xR_a y \Leftrightarrow f_a(x) = f_a(y)$.

By extending this definition to a subset of features such as $P \subseteq A$, x and y are indiscernible with respect to feature set P (denoted as $xR_P y$) if and only if $\forall x, y \in U, xR_P y \Leftrightarrow \forall (a \in P) f_a(x) = f_a(y)$.

The equivalence class of the element $x \in U$ with respect to feature set P, denoted as $IND(P)$ or $[x]_P$, is defined by $IND(P) = [x]_P = \{y | xR_P y\}$. This set consists of records whose values of P features are identical to those of x. For example, as shown in Table 5, options 15 and 753 are indiscernible with respect to features V1, V2, and V8, and they belong to the equivalence class P = {V1, V2, V8}.

Definition 6 (lower approximation with respect to the feature set): using the above symbols for each subset $X \subseteq U$, lower approximation with respect to the feature set P ($\underline{P}(X)$) is constructed as $\underline{P}(X) = \{x | [x]_P \subseteq X\}$. Let P and Q be two subsets of the feature set A. The positive region of Q with respect to P (denoted as $POS_P(Q)$) is defined as $POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}(X)$, where X is any partition of U with respect to Q (U/Q). In other words, the positive region of Q with respect to P represents the records whose values of Q features are determined by the values of P features.

Definition 7 (Dependency of features): the set of features, Q, is said to be completely dependent on the set of features, P, denoted as $P \Rightarrow Q$, if the value of each feature in Q can be determined using the values of the P features. Degree of dependency of the Q feature on P

($\gamma_P(Q)$): for $P, Q \subseteq A$, Q is called dependent on P with a degree of k ($(0 \leq k \leq 1)$) if $k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$.

This concept is represented by the symbol $P \Rightarrow_k Q$. If $k = 1$, Q is considered to be completely dependent on P; if $k < 1$, Q is partially dependent on P to a degree of k; and if $k = 0$, Q is not dependent on P (Jensen & Shen, 2001). Now, if Q is a subset of outcome features and P is a subset of information features, and the degree of dependency of Q on P is equal to one, the outcome of each record in the information table can be determined based on the values of the feature set P. By this definition, the set P is a reduct of features in the information table, and the degree of dependency of the outcomes on them is equal to the desired value (usually one).

In the theory of rough sets, the reduct with the least number of members is important. This reduct is called the minimum reduct. One way to obtain the minimum reduct is to calculate the dependency of the outcome set on all possible subsets of features. Any subset for which $\gamma(Q) = 1$ is a reduct, and the smallest subset with this feature will be the minimum reduct. However, this solution would not be suitable for large datasets because, assuming the number of informational features is n, there will be $2^n - 1$ subsets (excluding the empty set), and calculating the dependency of the outcomes on each of these subsets would be time consuming if the n value is large. The fast reduction algorithm is a heuristic algorithm that allows calculating a reduct close to the minimum without generating all possible subsets. The algorithm starts with an empty set and then, in a step-by-step approach, adds features that increase $\gamma_P(Q)$ to the feature set. This process continues until a desirable value close to 1 is achieved for the dataset. The algorithm is executed as follows:

1. $P \leftarrow \{\}$
2. Do
3. $T \leftarrow P$
4. $\forall x \in (A - P)$
5. If $\gamma_{p\cup\{x\}}(\mathrm{Q}) > \gamma_T(Q)$
6. $T \leftarrow P \cup \{x\}$
7. $P \leftarrow T$
8. Until $\gamma_P(Q) = \gamma_C(Q)$
9. Return P

In this algorithm, $\gamma_C(Q)$ is minimum value of outcome dependency on the determined features, which is determined at the beginning of the algorithm.

However, this method typically does not produce a minimum reduct; instead, it generates a reduct that is close to the minimum reduct (Swiniarski & Skowron, 2003). The variable precision rough set model used in this study is an extension of the rough set theory proposed by Ziarko (1993). The variable precision rough set model allows subjects to be classified with an error level lower than a predetermined threshold.

Let $X, Y \subseteq U$, the relative classification error ($C(X,Y)$) can be defined as $C(X,Y) = 1 - \frac{|X \cap Y|}{|X|}$, where $|X|$ represents the size of set X. Rough inclusion is achieved by allowing a specific level of error in the classification. By this definition, X is a β-subset of Y ($X \subseteq_\beta Y$) if and only if $X \subseteq_\beta Y \; iff \; C(X,Y) \leq \beta$ for $0 \leq \beta \leq 0.5$. This concept is referred to as rough inclusion. Using $\subseteq_\beta$ instead of ⊆, β-lower approximation of set X ($\underline{P}_\beta(X)$) can be defined as follows (Ziarko, 1993):

β-lower approximation of set X with respect to the equivalence relation P is defined as $\underline{P}_\beta(X) = \{x | [x]_P \subseteq_\beta X\} = \left\{x \middle| \frac{|X \cap [x]_P|}{|[x]_P|} \geq 1 - \beta\right\}$. Essentially, β-lower approximation of set X is a set of records whose equivalence class with respect to the set of features P is a β-subset of X.

The positive region of Q with respect to P in terms of β, denoted as $POS_{P,\beta}(Q)$, is defined as $POS_{P,\beta}(Q) = \cup_{X \in U/Q} \underline{P}_\beta(X)$ where P and Q are subsets of features. Using the above definitions, for example, assuming β=0.2, the positive region of set X with respect to P features will be the union of equivalence classes whose more than 80% members belong to set X. In this model, the dependency function denoted by $\gamma_{P,\beta}(Q)$ is calculated as $\gamma_{P,\beta}(Q) = \frac{|POS_{P,\beta}(Q)|}{|U|}$. In the following section, calculations related to computing $\gamma_{P,\beta}(Q)$ will be presented based on the given definitions.

Results

The second section discusses the formation of a new process or business. As noted in section 3-1, three decision points were identified during the process. In this section, we attempt to identify decision models at these three points using the algorithm for decision model inference and process model adaptation (Bazhenova & Weske, 2016), leveraging the theory of rough sets and the fast reduction algorithm. To run the fast reduction algorithm and evaluate the three decision points, the available data were divided as shown in Table 4.

Table 4. Number of data used to construct and evaluate the model

Decision point	Options (n)	Options used to construct	Options used for evaluation
dp_1	923	873	50
dp_2	409	369	40
dp_3	156	126	30

Furthermore, $\beta = 0.2$ and a desired degree of feature dependence on outcomes equal to 0.8 were assumed throughout this section.

First Decision Point: Decisions on the results of pre-feasibility studies

As discussed in section 2, each idea at the initial decision point is characterized by 13 features (derived from idea profile and results of pre-feasibility studies). Based on this and the decisions made by the portfolio committee, a decision log matrix (PIE) is formed. This matrix will be a 14×876 matrix. Table 4 presents a subset of this matrix.

Table 5. Subset of the decision matrix

Idea index	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	Result
15	2	1	1	2	2	2	4	2	3	1	2	2	2	Accepted
111	1	3	2	3	3	2	4	2	3	3	2	2	2	Rejected
463	3	2	3	3	1	3	4	1	2	2	4	1	1	Rejected
753	2	1	2	3	3	1	1	2	1	2	1	3	3	Rejected

Now, step by step, we will try to identify the effective factors in the decision-making process using a fast reduction algorithm.

Identifying the First Effective Critical Feature

To identify the first critical feature for the initial decision point by running the algorithm in the first phase, the results are presented in Table 6. Accordingly, the most important critical feature is the relevance of the idea to existing businesses."

Table 6. First run of the fast reduction algorithm for identifying the first effective factor

P	$\gamma_P(Q)$
V1	0.128
V2	0.102
<u>V3</u>	<u>0.212</u>
V4	0.202

P	$\gamma_P(Q)$
V5	0.112
V6	0.164
V7	0.089

P	$\gamma_P(Q)$
V8	0.111
V9	0.160
V10	0.180

P	$\gamma_P(Q)$
V11	0.111
V12	0.089
V13	0.100

Identifying Other Effective Features

By running the algorithm in 5 steps, the critical features and their degree of dependency are presented in Table 7. Accordingly, the critical features for the first decision point are identified as follows:

1. Relevance of the idea to existing businesses
2. Idea presenter
3. Expected customer acceptance
4. Number of competitors
5. Project implementation method

Table 7. Results of running the fast reduction algorithm for identifying effective features on decision-making in the 1st decision point

Step	Identified effective features	Degree of dependence
2	$V_1\} \text{ ‹} \{V_3$	0.355
3	{V1, V3, V10}	0.529
4	{V1, V3, V9, V10}	0.692
5	{V1, V3, V7, V9, V10}	0.801

Identifying the Decision Model

Using equivalence relationships in the last iteration of the fast reduction algorithm, inference rules are extracted and a critical feature tree is constructed based on these rules. The number of levels in the tree is determined by the fast reduction algorithm (in this case, 5). The top level of the tree partitions different decision outcomes by considering only the status of the first critical feature. The next level partitions different decision outcomes by considering the first and second features. Adding levels to the tree continues to the number of critical features. The partition precision of the outcomes at each level is equal to $\gamma_P(Q)$ identified features at the respective level. Figure 4 presents a part of the initial decision tree for the first decision point.

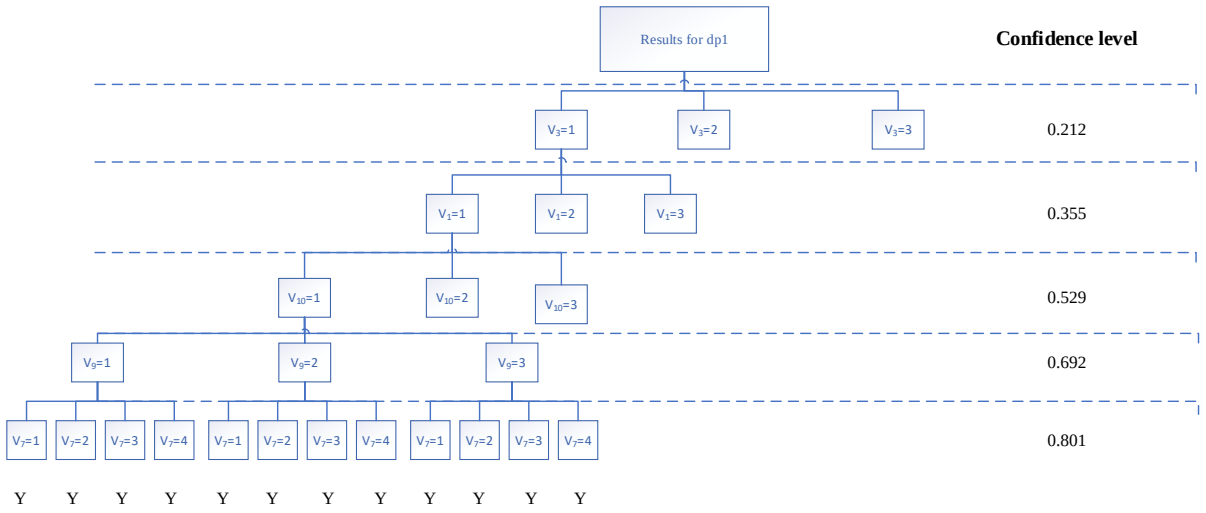

Figure 4. Part of the unpruned decision tree for the 1st decision point

After constructing the critical feature tree, the second step is to prune and finalize it. As shown in Figure 5, the critical feature tree can have one of the following conditions:

1. N: This means that the decision outcome for the conditions related to this leaf is negative at a significance level of 80%.
2. Y: This means that the decision outcome for the conditions related to this leaf is positive at a significant level of 80%.

3. NA: Among the available options (ideas in the PIE matrix), no option corresponds to the mentioned conditions on the leaf.
4. MC: Leaves with different decision outcomes and different ideas.

Regarding leaves with uncertain decision outcomes (NA and MC), the decision required (DR) label is attached in the decision tree. For the other leaves, a decision-making rule base is presented as a decision tree through aggregation and summarization. (Figure 5 presents a part of this decision tree.)

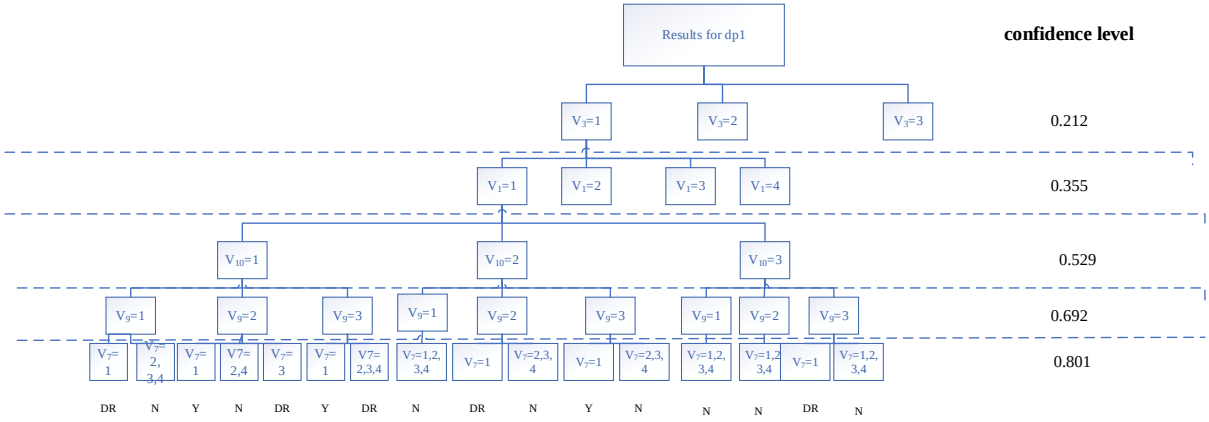

Figure 5. Part of the decision tree for the 1st decision point

As a result, out of 432 possible cases (equal to the product of the conditions of 5 identified features (3 × 4 × 3 × 3 × 4)), 37 leaves (approximately 8.5%) were labeled as decision required, and for the remaining leaves, the decision outcome was determined based on the available data.

Evaluating the Decision Model

After constructing the decision tree, it needs to be evaluated. For this purpose, test data are used. As noted, 50 data points were selected as test data from the available dataset in this phase. If the estimated and actual decision outcomes match an acceptable percentage of the test data (at least 1-β), the model is accepted; otherwise, it is rejected. Model rejection can be due to a lack of information consistency, which in turn indicates insufficient parameters for model construction and evaluation (Mansmann et al., 2007), which requires problem remodelling. In this problem, the decision outcomes matched the decision tree for 46 data points (equivalent to 92%), so the proposed decision model is considered acceptable.

Second Decision Point: Decisions on the results of pre-feasibility studies

By the steps mentioned regarding the data related to the second decision point, the results of the fast reduction algorithm are presented in Table 8.

Table 8. Results of the fast reduction algorithm for the 2nd decision point

Step	Identified critical features	Degree of dependence
1	{V17}	0.154
2	{V9, V17}	0.306
3	{ V9, V17, V21}	0.691
4	{V9, V16, V17, V21}	0.780
5	{V9, V16, V17, V21, V27}	0.808

After constructing and pruning the decision tree using 40 test data points, the outcomes matched the model for 92.5% of them, indicating acceptance of the model.

Third Decision Point: Decisions on Converting a Business Opportunity to a Project or Business Model

By following the steps outlined regarding the data related to the third decision point, the outcomes of the fast reduction algorithm are presented in Table 9. After constructing and pruning the decision tree using 30 test data points, the outcomes matched the model for 83.3% of them, indicating acceptance of the model.

Table 9. Results of the fast reduction algorithm for the 3rd decision point

Step	Identified critical features	Degree of dependence
1	{V27}	0.437
2	{V11, V27}	0.635
3	{V7, V11, V27}	0.786
4	{V7, V8, V11, V27}	0.865

Discussion

By checking the results of the fast reduction algorithm, the following results can be obtained:

- Critical decision features at this decision point are "Relevance of the idea to the existing business," "Source of the idea presenter," "Expected customer acceptance," "Number of competitors," and "Idea implementation." Based on the running features, decision rules for the decision point were also identified. According to the critical features, decision rules were identified based on the equivalence relationships of the last identified table, as follows:

"If the proposed idea is highly relevant to the existing business, the idea presenter is one of the employees, the expected customer acceptance is estimated to be high, and there are many competitors in the market. If the project type is EPC, the decision needs to be made by the portfolio committee. However, for other project execution models, the decision outcome is likely to be negative."

"If the proposed idea is highly relevant to the business and the idea presenter is also one of the owners, and if the expected customer acceptance is estimated to be high, regardless of

the implementation method and the number of competitors, the decision outcome is predicted to be positive."

- As noted earlier, almost 91.5% of different states related to critical feature states can be achieved using the decision model presented at this decision point. The remaining 8.5% had either multiple answers based on identified critical features (3.2%) or did not exist among options (5.3%). Confirming the identified rules will increase the average decision-making rate at this decision point by approximately 10 times, which is significant.

- In the second decision point, critical features "business opportunity assessment time," "number of competitors," "estimated maximum required capital," "estimated payback period," and "estimated net profit" were identified as critical features of the second decision point, and the decision tree was formed based on these features. Note that in the model presented at this decision point, only 176 leaves have a definite label (only 30.6%) considering the various states of critical features (number of decision tree leaves) – 4×4×4×3×3=576. This indicates a lack of information in the evaluation process. However, it should be noted that, according to the model described in the first decision point, for example, the fact that the expected customer acceptance is high in most ideas (69.1%) suggests that the expected payback period and, of course, the expected profit should be high. Considering that 49.3% of the accepted ideas are of the EPC type, where the price is paid by the employer, the payback period is expected to be very short. Therefore, most of the leaves with the NA label do not occur among the options. Evidence of this is the absence of this data among the test and model validation data. In the third decision point, four features, "business opportunity assessment time," "need for license," "project implementation form," and "idea novelty" were identified as critical features of the decision. At this decision point, data shortage is more pronounced than that at previous decision points.

- Identification of decision models at the second and third decision points by the portfolio committee decision makers leads to improved decision-making at previous decision points. For example, at the third decision point, the need for a license was identified as a critical feature, and most of the rejected options were ideas or opportunities that required a license. By publishing these results, this aspect should also be considered in previous decision points, which has been overlooked until now.

Conclusion

Based on the model presented in Bazhenova & Weske (2016) and using the theory of rough sets and the fast reduction algorithm, a step-by-step procedure was introduced to identify decision models at decision points. The algorithm was applied in one of the information technology companies to analyze decision points in the new product development process. By

identifying critical decision features, a decision tree was created for each decision point. It was observed that implementing this model and automating decision-making at one of the decision points resulted in a 90% reduction in the number of decisions and improved process efficiency. This undoubtedly contributed to the operational speed in this area.

However, it should be noted that in the abovementioned process, one of the critical features for the third decision point in the decision-making process was overlooked in the first and second decision points. Introducing the identified model for this point to decision makers can lead to the rejection of business opportunities that do not meet the desired criteria for this feature, thus reducing the workload associated with such opportunities.

A limitation of this study is the lack of sufficient data for analyzing decision-making, especially at the second and third decision points. It is natural that the usage and evaluation of the proposed model in large organizations, where the volume of available data in their decision logs exceeds that of the case study, will further demonstrate the strengths and weaknesses of the presented model. This can be suggested as an avenue for future research. In addition, applying the proposed model in other knowledge-intensive processes such as health or emergency management could be another potential area for future research.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The author's received no financial support for the research, authorship, and/or publication of this article.

References

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. *Mining text data*, 163-222.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational issues of big data and analytics in networked business. *MIS quarterly*, *40*(4).

Bazhenova, E., & Weske, M. (2016). Deriving decision models from process models by enhanced decision mining. In *Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–September 3, 2015, Revised Papers 14* (pp. 444-457). Springer International Publishing.

Bolisani, E., & Scarso, E. (1999). Information technology management: a knowledge-based perspective. *Technovation*, *19*(4), 209-217.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57.

De Leoni, M., & van der Aalst, W. M. (2013, March). Data-aware process mining: discovering decisions in processes using alignments. In *Proceedings of the 28th annual ACM symposium on applied computing* (pp. 1454-1461).

De Roock, E., & Martin, N. (2022). Process mining in healthcare–An updated perspective on the state of the art. *Journal of biomedical informatics*, *127*, 103995.

Ebrahimi Kordlar, A. & Mohammadi Shad, Z. (2014). Investigating the Relationship between Default Risk and Earning Response Coefficient (ERC). *Accounting and Auditing Review*, 21(1), 1-18. doi: 10.22059/acctgrev.2014.50780

Ebrahimi Kordlar, A., & Aarabi, M. J. (2010). Ownership concentration and earnings quality in the listed companies in tehran's stocks exchange. *Journal of financial accounting research*, 2(4), 95-110.

Fauzi, R., & Andreswari, R. (2022). Business process analysis of programmer job role in software development using process mining. *Procedia Computer Science*, *197*, 701-708.

Ghasemi, M., & Amyot, D. (2016). Process mining in healthcare: a systematised literature review. *International Journal of Electronic Healthcare*, *9*(1), 60-88.

Ghazal, M. A., Ibrahim, O., & Salama, M. A. (2017, November). Educational process mining: a systematic literature review. In *2017 European Conference on Electrical Engineering and Computer Science (EECS)* (pp. 198-203). IEEE.

Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388-423.

Gupta, B., Rawat, A., Jain, A., Arora, A., & Dhami, N. (2017). Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications*, *163*(8), 15-19.

Hasić, F., Serral, E., & Snoeck, M. (2020, March). Comparing BPMN to BPMN+ DMN for IoT process modelling: a case-based inquiry. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (pp. 53-60).

Heudecker, N. & J. Hare. 2016. Survey Analysis: Big Data Investments Begin Tapering in 2016. *Gartner Report*. Available at: https://www.gartner.com/en/documents/3446724 [Accessed 10.05. 2020].

Jensen R. and Shen Q. (2001) "A Rough Set Aided System for Sorting WWW Bookmarks", Proceedings of the First Asia-Pacific Conference on Web Intelligence: Research and Development, Springer-Verlag, London, UK

Jokonowo, B., Claes, J., Sarno, R., & Rochimah, S. (2018). Process mining in supply chains: a systematic literature review. *INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING*, *8*(6), 4626-4636.

Leymann, F., & Roller, D. (1999). *Production workflow: concepts and techniques*. Prentice Hall PTR.

Marjanovic, O., Skaf-Molli, H., Molli, P., & Godart, C. (2007, November). Collaborative practice-oriented business processes Creating a new case for business process management and CSCW synergy. In *2007 International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2007)* (pp. 448-455). IEEE.

Mazorodze, A. H., & Buckley, S. (2020). A review of knowledge transfer tools in knowledge-intensive organisations. *South African Journal of Information Management*, *22*(1), 1-6.

Mishra, A. K. (2019). Using Data Mining to Improve Healthcare Decision-Making: A Systematic Review. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *10*(1), 575-581.

Mohemad, R., Hamdan, A. R., Othman, Z. A., & Noor, N. M. M. (2010). Decision support systems (DSS) in construction tendering processes. International Journal of Computer Science Issues, 7, 35–45

Moreira, M. W., Rodrigues, J. J., Korotaev, V., Al-Muhtadi, J., & Kumar, N. (2019). A comprehensive review on smart decision support systems for health care. *IEEE Systems Journal*, *13*(3), 3536-3545.

Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., ... & Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, *127*, 103994.

Nikabadi, M. S., & Sepehrnia, A. (2019). The effect of knowledge-based information technology tools on the new product development processes in software companies. *International Journal of Business Innovation and Research*, *18*(1), 19-46.

Norozi, M., Piri, M., & Zahedi, M., (2026). Evaluating cybersecurity risks in IoT-enabled retail: A hybrid pythagorean Fuzzy-SWARA–ARTASI approach, *Industrial Management Journal*, 18(1), 29-61. https://doi.org10.22059/imj.2025.400267.1008259

Pawlak, Z. (1991) "Rough Sets: Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishing, Dordrecht

Pawlak, Z., Polkowski, L., & Skowron, A. (2001). Rough set theory. *KI*, *15*(3), 38-39.

Pedro, J., Brown, I., & Hart, M. (2019). Capabilities and Readiness for Big Data Analytics. *Procedia Computer Science*, *164*, 3-10.

Poppe, E., Pika, A., Wynn, M. T., Eden, R., Andrews, R., & ter Hofstede, A. H. (2021). Extracting Best-Practice Using Mixed-Methods: Insights and Recommendations from a Case Study in Insurance Claims Processing. *Business & Information Systems Engineering*, 1-15.

Portolani, P., Savoia, D., Ballarino, A., & Matteucci, M. (2023, May). A Novel Decision Mining Method Considering Multiple Model Paths. In *International Conference on Business Process Modeling, Development and Support* (pp. 79-87). Cham: Springer Nature Switzerland.

Reder, B., Freimar, A. J., Holzwart, G., Mauerer, J., Schonschek, O., & Schweizer, M. (2019). Studie Process Mining & RPA 2019. *Munich, Germany*.

Reichert, M., & Weber, B. (2012). *Enabling flexibility in process-aware information systems: challenges, methods, technologies*. Springer Science & Business Media.

Rozinat, A., & van der Aalst, W. M. (2006, September). Decision mining in ProM. In *International Conference on Business Process Management* (pp. 420-425). Springer, Berlin, Heidelberg.

Srivastava, S. (2021). Process mining techniques for detecting fraud in banks: A study. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(12), 3358-3375.

Swiniarski, R. W. and A. Skowron (2003) "Rough Set Methods in Feature Selection and Recognition", Pattern Recognition Letters, vol. 24, pp. 833-849

Urrea-Contreras, S. J., Flores-Rios, B. L., Astorga-Vargas, M. A., & Ibarra-Esquer, J. E. (2021, August). Process Mining Perspectives in Software Engineering: A Systematic Literature Review. In *2021 Mexican International Conference on Computer Science (ENC)* (pp. 1-8). IEEE.

Van der Aalst, W. M. (2022). Process mining: a 360 degree overview. In *Process Mining Handbook* (pp. 3-34). Springer, Cham.

Van Dongen, B. F., De Medeiros, A. A., & Wen, L. (2009). Process mining: Overview and outlook of petri net discovery algorithms. In *transactions on petri nets and other models of concurrency II* (pp. 225-242). Springer, Berlin, Heidelberg.

Wang, J., Wong, R. K., Ding, J., Guo, Q., & Wen, L. (2012). Efficient selection of process mining algorithms. *IEEE Transactions on Services Computing*, *6*(4), 484-496.

Weber, B., Reichert, M., Rinderle-Ma, S., & Wild, W. (2009). Providing integrated life cycle support in process-aware information systems. *International Journal of Cooperative Information Systems*, *18*(01), 115-165.

Weske, M. (2019). *Business Process Management: Concepts, Languages, Architectures*. Springer.

Yin, D., Dong, L., Cheng, H., Liu, X., Chang, K. W., Wei, F., & Gao, J. (2022). A survey of knowledge-intensive nlp with pre-trained language models. *arXiv preprint arXiv:2202.08772*.

Zhu, J., He, P., Fu, Q., Zhang, H., Lyu, M. R., & Zhang, D. (2015, May). Learning to log: Helping developers make informed logging decisions. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering* (Vol. 1, pp. 415-425). IEEE.

Ziarko W. (1993) "Variable Precision Rough Set Model", Journal of Computer and System Sciences, vol. 46, pp. 44-54.

Bibliographic information of this paper for citing:

Chehrehpak, Mehri; Toloui Ashlaghi, Abbas & Mohammadkhani, Kamran (2026). A Rough Set–Based Process Mining Approach for Knowledge Intensive Processes: Insights from a New Product Development Case Study. *Journal of Information Technology Management*, 18 (3), 25-46. https://doi.org/10.22059/jitm.2026.402437.4270