IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4291~4300
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4291-4300

Developing standard criteria for robotic process automation candidate process selection

Neelam Yadav[1,2], Supriya P. Panda[1]
[1]Department of Computer Science and Engineering, School of Engineering and Technology, Manav Rachna International Institute of Research and Studies, Faridabad, India
[2]Department of Computer Applications, ABES Engineering College, Ghaziabad, India

Article Info

Article history:

Received Mar 5, 2024
Revised Apr 21, 2024
Accepted Jun 1, 2024

Keywords:

Business process prioritization
Candidate process selection
machine learning
Naïve bayes
Robotic process automation
Robotic process mining
Software bots

ABSTRACT

Robotic process automation (RPA) is a cutting-edge technology that provides software robots to repeat and mimic the repeatable tasks that a human user earlier performed. The use of software robots is encouraging because of their cost efficiency and easy implementation. Selecting and prioritizing a candidate process for automation is always challenging as all the business processes in an organization are not equally suitable for RPA implementation. Various studies have highlighted several criteria found in the literature for determining, prioritising, and selecting a business process for RPA. Nevertheless, there are no set standards for evaluating and analyzing a certain process or its tasks to determine whether they may be automated to use RPA. This paper aims to develop standard criteria and propose a consistent model to select and prioritize candidate process for RPA projects. To assess these criteria's applicability in the context of RPA, surveys among subject matter experts (SMEs) are used to validate them. Principal component analysis (PCA) and correlation are used to identify the top 20 criteria. Naïve Bayes algorithm is applied on the collected data for decision-making. The developed multi-criteria model exhibits strong precision and recall measures, with training and validation accuracy of 96% and 90%, respectively.

Corresponding Author:

Neelam Yadav
Department of Computer Science and Engineering, School of Engineering and Technology
Manav Rachna International Institute of Research and Studies
Faridabad, Haryana, India
Email: mail2neelam.sanskriti@gmail.com

1. INTRODUCTION

Industry 4.0 heralds a new era where intelligent human intervention is essential for highly intellectual and cognitive tasks, reshaping the landscape of advanced manufacturing and technological innovation [1]. Industrial robots have revolutionized industries by replacing human labor in structured activities. Enterprises are continually driven to optimize and automate their manual processes, leading to a crucial need for digitizing and automating monotonous and repetitive business processes [2]. Robotic process automation (RPA) has emerged as a pivotal technology, introducing virtual workforces in the form of software bots [3]. RPA streamlines business processes, enhancing efficiency and cutting operational costs by reducing full-time equivalents (FTEs) [4]. Leveraging graphical user interfaces (GUIs), RPA automates and interconnects systems, closely emulating human-computer interactions (HCIs) without necessitating IT expertise for implementation [5]. RPA's advantages include cost reduction, increased agility, improved quality, and enhanced customer satisfaction, all while being minimally intrusive [6]. It seamlessly integrates with customer relationship management (CRM) and enterprise resource planning (ERP) systems, enabling organizations to focus on more strategic endeavors [7]. By mimicking tasks previously performed by humans, RPA frees up human resources for more cognitive work, fostering employee development, customer satisfaction, and skill acquisition [8].

However, many large organizations still have numerous non-automated business processes, prompting them to consider automating these processes with RPA, which promises significant and rapid returns [9]. RPA projects follow the software development life cycle (SDLC), involving analysis, design, development, implementation, and testing phases. Traditional analysis methods such as interviews with subject matter experts (SMEs) and process documentation analysis are prone to biases and a lack of documentation [10]. This lack of standardized approaches for prioritizing and selecting processes for automation with RPA poses a risk of economic disinvestment for prospective adopters [11]. Academia and industry are therefore in pursuit of a multi-criteria model for successful RPA implementation [12]. This paper aims to address this gap by proposing standardized criteria for process selection, enabling practitioners and researchers to identify suitable business processes for RPA implementation .

RPA offers software solutions for automating repetitive and labor-intensive tasks within business processes. Through software scripts, digital bots can access uniform resource locators (URLs) and system applications to view, extract, and process data [13]. Process discovery techniques, such as extracting user interface (UI) logs, facilitate RPA implementation [14]. Existing case studies in various industries have primarily focused on volume and complexity when selecting processes for RPA, but the lack of standardized criteria poses challenges [15]. Without well-defined criteria, there is an increased risk of project failure and hindrance in realizing RPA benefits [16]. Therefore, standardized process selection methodologies are crucial for maximizing the potential benefits of RPA and mitigating risks associated with process automation [17]. This research addresses the challenge of identifying multiple criteria for selecting candidate processes in RPA projects. The following conducts a literature review to confirm these as areas requiring further research. The subsequent section also emphasizes that the multi-criteria model is trained and validated using a machine learning methodology. In contrast to earlier studies, our training model achieves an overall accuracy of 96%, indicating a 6% enhancement.

– Literature review and research gaps

To date, researchers have addressed the challenges by providing suitable measures and criteria for picking RPA suitable processes [16]. The literature review focuses on three widely recognized approaches: one centered on process quality and characteristics, another reliant on interviews, questionnaires, and discussions [17], [18], and the third rooted in process mining. Attributes like stability, error rate, maturity, standardization, input and maturity are accessed in process quality criteria as discussed in Table 1. The robotic process mining techniques are quite relying and promising to increase RPA appropriateness and suitability [19]. Process mining works on event logs or UI logs. The UI logs are trapped and recorded while a human user is interacting with the underlying system. These UI logs are used to make process maps. The process maps are further analyzed by discovering automatable process. Collection of quality UI logs for process mining algorithms are quite difficult. Lack of quality UI logs for each un-automated process, makes this method unsuitable for large or medium sized organizations. A popular set of indications, including rule-driven, repetitive, data-intensive, high compliance, and validations to choose the right candidate, was introduced by Lacity and Willcocks [7]. Academics and industry professionals rely more on conducting interviews with industry experts, utilizing questionnaires and employing discussion-based methods. These techniques are very often used by RPA experts in the analysis phase of RPA projects. Many authors contribute valuable insights into the criteria shaping the process selection landscape. The best suited characteristics of a process for RPA are rules-based, low complexity, repetitive, voluminous transactions (higher number of occurrences), digitization, structured, matured, standardized, documented, and interaction with multiple systems [18], [20], [21]. A set of quantitative measures based on process based characteristics were proposed including automation rate, execution frequency, standardization and maturity [13]. It has been found that business processes with a high transaction volume, process maturity, and business rules are appropriate for RPA and can result in great success [12], [21], [22]. Furthermore, business processes with low complexity, error proneness, low cognitive requirements, and high workload are found to be better candidates for RPA bot installation [23].

Criteria including duration [2], [24], execution time [13], [21], digitization [23], [25], risk proneness [23], limited human intervention and low cognitive requirements [26] are not very promptly discussed in literature. These criteria are often overlooked or given less importance in the existing literature on the subject. However, they play a crucial role in determining the effectiveness and efficiency of various processes and systems. This is the reason duration and cognitive requirements are included in our study. It is essential for future research to delve deeper into these aspects to gain a comprehensive understanding of their impact on different domains. While these connected works offer insights on evaluating a candidate business process's RPA applicability, a systematic procedure is lacking.

– Research targets

As such, it is essential to determine an appropriate early procedure in business analysis, to ascertain the candidacy. Some of the seventeen-identified criteria, which form the focal points of discussion, are described in Table 1. While certain criteria are emphasized consistently across multiple studies, nuances, and variations in their importance emerge, highlighting the need for a comprehensive understanding of the diverse factors influencing the process selection process. These nuances and variations can be attributed to factors such as organizational culture, industry context, and individual preferences. Therefore, it is crucial for decision-makers to consider these diverse factors when making process selection decisions to achieve optimal outcomes. This research addresses the challenge of identifying multiple criteria for selecting candidate processes in RPA projects. However, a very little work has been done on prioritization and selection of automation of processes. To access the automation potential of business processes both practitioners and researchers are focusing on a one step solution. So, this research along with addressing the challenge of identifying multiple criteria for selecting candidate processes in RPA projects also aims to develop a multi-criteria model for organizations to better select and prioritize business processes that are best suited for RPA.

Table 1. The seventeen-identified criteria from literature

Process Selection Criteria	[2]	[12]	[13]	[18]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]
Clear costs						●	●	●		●			
Stability	●		●			●	●	●		●	●		●
Voluminous transactions		●		●	●	●	●	●	●	●		●	●
Multiple systems		●		●		●	●	●	●	●	●		●
Rules based		●		●	●	●	●	●	●	●	●	●	●
Standardization	●	●	●	●	●		●	●	●				
Error prone					●					●	●		
Few exceptions					●	●	●			●			
Execution frequency	●		●					●		●			
Automation rate	●		●		●								
Process complexity		●		●	●				●				●
No. of FTEs					●		●						
Repetitive	●				●				●			●	
Structured digital data	●			●	●							●	
Maturity		●		●			●					●	
Limited human intervention							●						
Structured data	●						●					●	

2. PROPOSED METHOD

It's crucial to understand the characteristics of the process and the significance of each element in achieving success. In this study, a three-step multi-criteria model is proposed to assess the suitability of a business process for RPA. This model involves five steps for identifying and categorizing process suitability for RPA. Figure 1 shows the research methodology comprising the multi-criteria process selection model for RPA. The research methodology comprises eight steps:

Step 1. Conduct a literature review to identify RPA process suitability criteria.
Step 2. Prepare a survey questionnaire for SMEs, including HEI-related processes.
Step 3. Distribute the questionnaire via social media and emails, instructing SMEs to select one process.
Step 4. Collect and analyse SME responses to identify trends in process suitability for RPA.
Step 5. Principal component analysis (PCA) and correlation analysis (CA) criteria selection technique are used to finalize the 22 criteria.
Step 6. Develop multi-criteria process selection models based on literature and survey findings.
Step 7. Validate and refine models through SME and industry expert consultations.
Step 8. Document findings and models in a detailed report and present outcomes to HEI administrators, RPA practitioners, and researchers.

The entire questionnaire was developed using the established criteria from the literature review. This questionnaire includes the processes like examination, placement, admission, attendance tracking, course planning and scheduling, and lesson plan creation. The SMEs are highly skilled academicians (Professors, Associate Professors, Assistant Professors, and Research Scholars) from higher education institutions (HEIs) and the business world (RPA specialists) as shown in Figure 2. SMEs are required to select one of these processes and answer the entire questionnaire with that specific process in mind. A total of 1,007 records were retrieved through the questionnaire. 22 different criteria were identified from both the approaches (literature review and questionnaire) that could impact the decision for RPA suitability.

This multi criteria model comprises three sets of criteria including 22 overall criteria for process selection and prioritization. The discussion below covers three sets of criteria: process characteristics-based, commercial impact-based criteria, and robotic process mining-based criteria shown in Figure 1 model, each essential for evaluating process automation eligibility. The first set of criteria (process characteristics-based criteria) consists of 12 criteria, which is based on the behavior and characteristics of a process. The second set also consists of 8 criteria that evaluates and determines the commercial impact of RPA implementation in a process. The third set comprises two criteria based on robotic process mining technique. Among 22 criteria, significant 20 criteria are included in a questionnaire. To access a process for automation potential, a process needs to be evaluated against each criterion. The total score of the assessing values will decide suitability of a process for RPA.

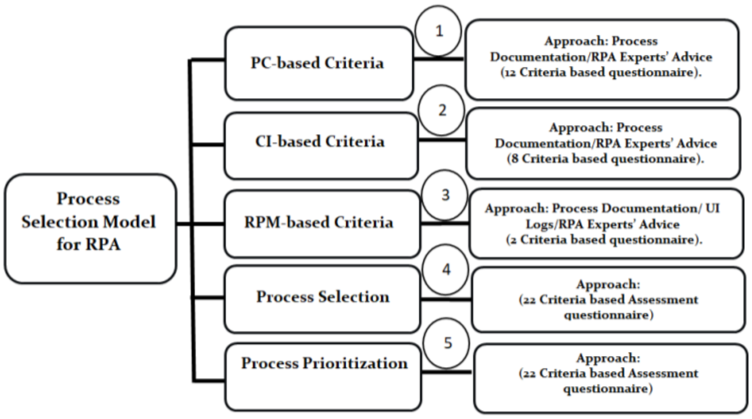


Figure 1. Multi-criteria model for RPA process selection, source: self

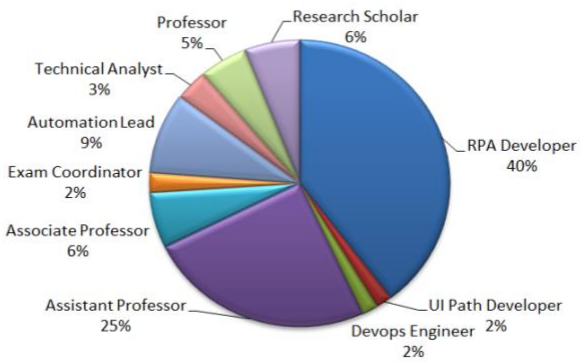


Figure 2. Demographic analysis of SMEs

2.1. Process characteristics-based criteria

To assess these process selection criteria several behaviour aspects of a process need to be assessed. Like to check whether a process is repetitive and rules based. One has to count the no of repetitive tasks. As RPA can only be beneficial, if a process has repetitive and monotonous tasks. Is there any human intervention while performing the tasks in a process? Is the process consists of well-defined rules or not. A voluminous process, if gets automated makes it more cost-effective. As RPA is a costly solution so it is always recommended that the process that undergoes for RPA has to be standardized, structured and error-free. If the rules of a process often change, then that process are least suitable for RPA. So, it is very necessary to ask how long a particular process gets executed without any changes and error. A stable, mature and standardized process is most suitable for automaton. So, a count, or a scale or any weighing factor should be provided to get the final score for all the process characteristics-based criteria as discussed in Table 2.

2.2. Commercial impact-based criteria

The second set of criteria is referred to as CI based criteria as discussed in Table 3. This criterion can specify how much gain an organization will receive, if its process is automated through RPA. These criteria cannot only be answered by stating yes or no. It also includes interval or ratio values. For instance, the criteria 'multiple systems involved' cannot be answered by yes or no. Rather it can be answered with a constant number like, no of multiple systems or applications are involved to complete a given tasks in a process. So, if a process involves interaction with multiple systems has high chances of repetition of data like copying from one system and pasting it in multiple applications or systems, requires manual effort and hence, making it more feasible for RPA. All the commercial impact-based criteria have a high impact in terms of increasing business value, because if a process is automated by assessing these criteria, the productivity will be increased thereby reducing the FTEs and it will increase the business values of a process. That’s why all the criteria are accessed so as to increase the commercial impact on a process.

Table 2. Process characteristics-based criteria

Assessment criteria
Repetitiveness (REP)
Rule based (RULE)
Process Stability (P_STAB)
Maturity (P_MAT)
Standardization (STND)
Error-proneness (EP)
Voluminous (VOL)
Structured (STRUC)
Scalability (SCALE)
Cognitive (COG)
Lifetime of a process (LP)
No. of hours (HOURS)

Table 3. Commercial impact-based criteria

Assessment criteria
Data accessibility (DA)
Manual effort and time savings (ME)
Crucial business process (CBPS)
Multiple systems involved (MSI)
Human intervention (HI)
Adherence and compliance (AC)
Enhanced productivity (EP)
Duration (DUR)

2.3. Robotic process mining-based criteria

Process mining is a technique to mine a process to extract important insights about the real enactment of a business process. Process Mining works on event logs or UI logs and produce business process models that describe the behavior of a process. To produce an accurate process model, the UI logs should be very accurate so that the task can be replicated by following the user interactions recorded by the UI logs. A UI Action logger [30] will be required to store the UI logs for automated discover of process models. These process models are further used to prioritize and select a business process for automation via RPA. Process documentation will also help in prioritization of candidate processes for RPA implementation. So, the availability of high-quality UI logs and Textual process documentation as discussed in Table 4 are required to access the process against these criteria. UI logs and process documentation may not be maintained for many traditional processes; hence robotic process mining-based criteria are not included in the survey questionnaire.

Table 4. Robotic process mining-based criteria

Assessment criteria
Process documentation (PD)
UI logs (UIL)

3. METHOD

3.1. Data preprocessing

The survey questionnaire distributed across the industry yielded a collection of 1,007 records. These records contain responses comprising both ordinal and categorical data. Subsequently, the responses underwent a cleaning process and were transformed into a two-class decision-based predictive algorithm. This algorithm assigns labels of "yes" or "no" (where "yes" is represented by 1 and "no" by 0), recorded under the column titled "candidate business process selection (CBPS)". As described in Table 5, nominal criteria such as REP, RULE, P_STAB, P_MAT, and STND have response options of either "yes" or "no". Conversely, criteria like

VOL, LP, and HOURS necessitate analysis before being converted into binary values of 0 and 1. For example, HOURS (number of hours) is evaluated based on the average duration of a user's engagement in the process. Initially, the inputs for this criterion exhibited variations, including outliers and discrepancies, such as some users providing input in terms of total hours per week (e.g., 19 hours per week). These inputs underwent a cleaning process and were subsequently converted into binary values of 0 and 1. If the number of hours increased by 5, it was assumed to be automatable and therefore assigned a value of '1'.

Table 5. Finalized process assessment criteria for process selection and prioritization in RPA

Criteria type	Criteria name	Description	Score value	Total score
Process characteristics-based characteristics criteria	Repetitiveness (REP)	Are there tasks that you perform regularly, involving the same set of actions?	Yes:1 No: 0	1
	Rule-based nature (RULE)	Are there tasks or processes in your role that follow clear and well-defined rules or guidelines, making them suitable candidates for automation through rule-based decision-making?	Yes:1 No: 0	1
	Process Stability (P_STAB)	How often do the processes you're involved in undergo changes or modifications?	Yes:1 No: 0	1
	Process Maturity (P_MAT)	Are the administrative processes you're involved in well-documented, standardized, and actively improved?	Yes:1 No: 0	1
	Standardization (STND)	Is the process you're evaluating for automation well-documented and standardized, with clear guidelines and procedures that are consistently followed?	Yes:1 No: 0	1
	Error-Prone Tasks (ErrorP)	Are there tasks within your responsibilities that are prone to errors due to manual intervention?	Error-Prone: 0 Error Free: 5	1
	Voluminous (VOL)	How would you describe the volume of work associated with your current tasks?	Daily: 1 Weekly: 5 Monthly: 2	1
	Structured (STRUC)	Are the processes you work with well-defined and structured?	If yes then score is 1 otherwise 0.	1
	Scalability (SCALE)	Can the volume of work associated with your tasks be easily scaled up or down based on changing demands?	If yes then score is 1 otherwise 0.	1
	Cognitive (COG)	Does the process you are involved in require complex decision-making, critical thinking, or creative problem-solving that involves a deep understanding of context and nuanced judgment? Yes or No.	If yes then score is 1 otherwise 0.	1
	Lifetime of a process (LP)	How long do you anticipate the process you're evaluating for automation to remain relevant and necessary within the organization's operations? Please provide an estimate in terms of years.	1 year: 0 2 years: 5 >=5 year: 1	1
	No. of hours (HOURS)	What is the average duration of the process you are currently involved in? Please provide an estimate in terms of hours or days.	No. of hours <5:0 No. of hours>=5:1	
Commercial impact-based criteria	Data accessibility (DA)	Is the data required for the process readily accessible and available in a digital format, making it easy to retrieve and use for automation purposes?	If yes then score is 1 otherwise 0.	1
	Manual effort and time savings (ME)	Are there specific tasks in your daily responsibilities that you believe could benefit from automation, resulting in reduced manual effort and time savings?	If yes then score is 1 otherwise 0.	1
	Crucial Business Process (CBPS)	Is the process you're considering for automation essential to the core operations and functions of the organization?	If yes then score is 1 otherwise 0.	1
	Multiple Systems Involved (MSI)	How much of the process requires human interaction with multiple systems or applications?	If >3 then 1 otherwise 0	1
	Human Intervention (HI)	Can the tasks you perform be executed without any human intervention?	If yes then score is 1 otherwise 0.	1
	Adherence and Compliance (AC)	Could implementing RPA in the process help enhance adherence to regulatory standards, internal policies, and best practices, thereby increasing compliance levels?	If yes then score is 1 otherwise 0.	1
	Enhanced productivity (EP)	Do you believe that implementing RPA in the process would lead to an increase in overall productivity, allowing tasks to be completed more efficiently and effectively?	If yes then score is 1 otherwise 0.	1
	Duration (DUR)	How long has this process been in operation within the organization? (No. Of Years)	>=5 Years: 1 <5: 0	1
Robotic process mining-based criteria	UI Logs	Whether the UI logs are available for the concerned process.	Logs Available: 1 Otherwise: 0	1
	Process Documentation	Completeness of Process Documentation.	Documentation Available: 1 Otherwise: 0	1
			Total	22

3.2. Criteria selection

Following the data cleaning and preprocessing stage, both PCA and CA were conducted for all 20 criteria, as detailed in Figures 3 and 4. Each criterion exhibited a positive correlation with CBPS. Notably, human intervention (HI), duration (DUR), and crucial business process (CBPS) demonstrated correlations ≥0.5 with CBPS. Furthermore, human intervention (HI), duration (DUR), multiple system interaction (MSI), and process maturity (P_MAT) collectively attained the highest rank (0.835) in the PCA calculations. Repetitiveness (REP), RULE, and VOL, alongside HI and DUR, emerged as the top-ranked criteria. Consequently, out of the 20 criteria examined, 17 displayed a significant impact on candidate business selection, as illustrated in Figure 4. Attributes such as HI, DUR, MSI, P_MAT, REP, RULE, VOL, EP, ME, HOURS, SCALE, P_STAB, AC, ErrorP, DA, STRUC, and STND were identified as highly ranked features in PCA criteria selection. Conversely, criteria such as lifetime of a process (LP), Cognitive (COG), and data accessibility (DA) exhibited minimal or negligible impact on identifying processes suitable for RPA. Nevertheless, given the relative novelty of RPA and the resulting lack of awareness within the education domain, capturing information regarding these criteria remains crucial for RPA projects. After PCA and CA, the finalized criteria for selecting and prioritizing a process are discussed in Table 5. Questions asked in the survey questionnaire and the score value of criteria determines the total score for a particular criterion. A threshold value of '>=18' is identified as 'Automotable' by the supervised naïve Bayes machine learning model.

	CBPS	REP	RULE	P_STAB	P_MAT	STND	ErrorP	VOL	STRUC	SCALE	COG	LP	HOURS	DA	ME	CBP	MSI	HI	AC	EP	DUR
CBPS	1	0.22	0.22	0.22	0.22	0.13	0.18	0.21	0.2	0.29	0.33	0.29	0.13	0.18	0.21	0.4	0.29	0.59	0.06	0.15	0.58
REP	0.22	1	1	0.04	0.03	0.03	0.14	-0.07	0.05	0.01	0.05	0.12	0.01	0	0.03	0.08	0.16	0	-0.02	-0.04	0
RULE	0.22	1	1	0.04	0.03	0.03	0.14	-0.07	0.05	0.01	0.05	0.12	0.01	0	0.03	0.08	0.16	0	-0.02	-0.04	0
P_STAB	0.22	0.04	0.04	1	0.45	-0.02	-0.01	0.16	0.14	0.05	0.21	0.03	-0.1	0.06	0.12	0.05	0.18	0.09	0.02	0.11	0.05
P_MAT	0.22	0.03	0.03	0.45	1	-0.03	0.04	0.11	0.16	0.04	0.23	0.14	-0.06	0.22	0.24	-0.04	0.2	0.09	0.04	0.21	0.08
STND	0.13	0.03	0.03	-0.02	-0.03	1	0.05	0.09	0.02	-0.03	-0.01	-0.06	0.11	0	0.02	0.06	0.01	0.04	0.1	0.07	0.03
ErrorP	0.18	0.14	0.14	-0.01	0.04	0.05	1	-0.02	0.01	-0.13	-0.04	0.05	0.02	0.14	0.1	0.11	0.11	0.09	-0.03	0.1	0.08
VOL	0.21	-0.07	-0.07	0.16	0.11	0.09	-0.02	1	0.08	-0.09	0.12	-0.03	-0.24	0.03	0.01	0	-0.01	0.1	0.25	0.08	0.1
STRUC	0.2	0.05	0.05	0.14	0.16	0.02	0.01	0.08	1	0.19	0.12	0.08	-0.01	0.04	0.11	0.08	0.12	0.02	0.05	0.04	0.02
SCALE	0.29	0.01	0.01	0.05	0.04	-0.03	-0.13	-0.09	0.19	1	0.05	0.18	0.09	-0.02	0.13	0.27	0.21	0.07	-0.08	0.09	0.07
COG	0.33	0.05	0.05	0.21	0.23	-0.01	-0.04	0.12	0.12	0.05	1	-0.04	-0.03	0.01	0.06	0.09	0.06	0.19	0.04	0.07	0.17
LP	0.29	0.12	0.12	0.03	0.14	-0.06	0.05	-0.03	0.08	0.18	-0.04	1	0.07	0.2	0.11	0.23	0.21	0.04	-0.02	0.05	0.04
HOURS	0.13	0.01	0.01	-0.1	-0.06	0.11	0.02	-0.24	-0.01	0.09	-0.03	0.07	1	0.02	0.03	0.03	0	0.04	-0.04	0.04	0.05
DA	0.18	0	0	0.06	0.22	0	0.14	0.03	0.04	-0.02	0.01	0.2	0.02	1	0.37	-0.02	0.16	0.1	0	0.29	0.11
ME	0.21	0.03	0.03	0.12	0.24	0.02	0.1	0.01	0.11	0.13	0.06	0.11	0.03	0.37	1	0.05	0.16	0.11	0.02	0.4	0.1
CBP	0.4	0.08	0.08	0.05	-0.04	0.06	0.11	0	0.08	0.27	0.09	0.23	0.03	-0.02	0.05	1	0.14	0.18	0	-0.06	0.17
MSI	0.29	0.16	0.16	0.18	0.2	0.01	0.11	-0.01	0.12	0.21	0.06	0.21	0	0.16	0.16	0.14	1	0.08	0	0.2	0.07
HI	0.59	0	0	0.09	0.09	0.04	0.09	0.1	0.02	0.07	0.19	0.04	0.04	0.1	0.11	0.18	0.08	1	0.04	0.05	0.96
AC	0.06	-0.02	-0.02	0.02	0.04	0.1	-0.03	0.25	0.05	-0.08	0.04	-0.02	-0.04	0	0.02	0	0	0.04	1	0.04	0.02
EP	0.15	-0.04	-0.04	0.11	0.21	0.07	0.1	0.08	0.04	0.09	0.07	0.05	0.04	0.29	0.4	-0.06	0.2	0.05	0.04	1	0.03
DUR	0.58	0	0	0.05	0.08	0.03	0.08	0.1	0.02	0.07	0.17	0.04	0.05	0.11	0.1	0.17	0.07	0.96	0.02	0.03	1

Figure 3. Correlation matrix among 20 Criteria, source: WEKA 3.8.6

```
Ranked attributes:
 0.8395   1 -0.467CBPS-0.357HI-0.348DUR-0.243MSI-0.233P_MAT...
 0.7388   2 -0.603REP-0.603RULE+0.247HI+0.247DUR+0.209VOL...
 0.6481   3 0.399DUR+0.388HI-0.37EP-0.363P_MAT-0.33ME...
 0.5722   4 -0.43VOL+0.376HOURS+0.319SCALE-0.291P_STAB-0.284AC...
 0.5031   5 0.467SCALE-0.363ErrorP-0.338DA+0.311STRUC+0.292CBP...
 0.4447   6 0.541STND+0.507AC+0.335VOL+0.244CBP-0.241P_MAT...
 0.3927   7 -0.533HOURS+0.412LP-0.392STND-0.302COG+0.241CBP...
 0.3464   8 0.623ErrorP+0.291P_STAB+0.284STND-0.282AC+0.266CBP...
 0.3031   9 0.499LP+0.466HOURS+0.335AC+0.289P_MAT-0.249EP...
 0.2612  10 0.77 STRUC-0.343MSI-0.267STND+0.245ErrorP-0.201EP...
 0.2217  11 -0.657COG+0.357MSI-0.33CBP+0.292STRUC+0.249STND...
 0.1834  12 0.502MSI+0.498AC-0.423STND+0.288ErrorP-0.259DA...
 0.1488  13 0.424MSI+0.422COG-0.378ME-0.365P_STAB-0.335CBP...
 0.1165  14 0.564VOL+0.363HOURS+0.326EP-0.281STND-0.258AC...
 0.0879  15 0.536DA-0.446LP-0.359EP+0.274HOURS+0.266VOL...
 0.0613  16 -0.642SCALE-0.313ErrorP+0.301CBP-0.272P_MAT-0.266DA...
 0.0362  17 -0.612ME+0.494EP+0.341DA-0.252VOL+0.214P_STAB...
```

Figure 4. PCA for criteria selection, source: WEKA

3.3. Supervised classification model development

To maintain simplicity, the supervised classification model is chosen from multiclass classification to determine whether a proposed business process is suitable for RPA. The naive Bayes [31] method is one of the most widely used classification algorithms in data mining and machine learning. It is predicated on the Bayes theorem, which says that conditional probabilities can be used to calculate the likelihood of an event given the occurrence of another specific event. The classification model is applied using the WEKA tool [32]. The collected dataset is divided in to two categories training and testing data with a percentage split of 75% and 25%, respectively. The training dataset has 747 records, whereas the testing dataset has 256 records. The nominal attribute has two class labels (automotable and not automotable) to find out suitability of a particular process for RPA.

4. RESULTS AND DISCUSSION

This research presented a multi-criteria decision-making model to assist academics and RPA experts in selecting and prioritizing tasks or routines inside business processes for RPA adoption. This study set out to identify 22 unique criteria for evaluation of processes for RPA implementation. But previous studies have only examined a limited set of criteria. The characteristics of three distinct process selection approaches are combined in this study. The first is based on robotic process mining, the second on process characteristics, and the third is based on interviews. Prior research has not included a single set of unified standards that are coupled with robotic process mining-based standards. Our method aims to minimize the time and effort needed to find good candidates for RPA. The criteria are finalized and then compiled into a questionnaire that is sent to SMEs for validation. 1,007 records are gathered from the questionnaire used for the survey.

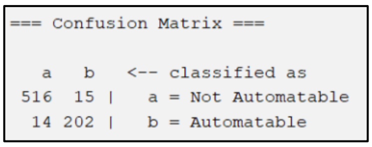
The multi-criteria model is trained and validated using a machine learning technique. Compared to the prior studies, our training model has an overall accuracy of 96%, which is 6% higher [29]. For this study, 1,007 professionals have provided insightful suggestions. As RPA is an expensive solution, our research indicates that additional criteria should be included when evaluating a process for RPA. For this conclusion, a process in its entirety is required. The suggested approach might be useful even in cases where traditional processes do not have UI or event logs, which were previously required for process selection. The results of training and validation model are discussed in subsequent sections.

4.1. Training model evaluation

The naïve Bayes algorithm evaluates the accuracy of the trained classification model. Among the selected 20 criteria, a threshold value ">=18" is chosen to determine if a process is automotable or not. The class label "automotable" is applied to every record in the dataset that satisfies this threshold value; otherwise, the class label "not automatable" is applied. Since naïve Bayes is a supervised machine learning approach, the training model's records must be labeled in order for the testing model to easily validate them. In machine learning, metrics like accuracy, precision, and recall assess different aspects of model quality. Precision measures prediction accuracy for the target class, while recall indicates the percentage of accurate results produced by the model. The F score, a weighted average of precision and recall ranging from 0 to 1, reflects overall performance. Table 6 presents the outcomes of this naïve Bayes classification model. The model has a 96% overall accuracy with good precision and recall measures as shown in Table 6. A higher F1- score specifies good accuracy of the training model. The confusion matrix depicted in Figure 5 has a true positive (TP) rate as 96%. It means 516 processes are correctly identified and 14 are wrongly identified by this model.

Table 6. Metrics' results of multi criteria training model, source: WEKA

Accuracy of the classification training model	Precision	Recall	F-Score	TP Rate	FP Rate
96%	0.960	0.960	0.960	0.960	0.058



```
=== Confusion Matrix ===

   a   b   <-- classified as
 516  15 |   a = Not Automatable
  14 202 |   b = Automatable
```

Figure 5. Confusion matrix of multi criteria training model, source: WEKA

4.2. Validation of the multi criteria model

The proposed multi-criteria model is validated with 256 records collected from the survey. These records are tested against the training model with naïve Bayes machine learning algorithm. Table 7 presents the results of validation of the trained multi-criteria model. It can be seen that the overall testing accuracy of the ML model is 90% with upright precision and recall measures. The validation model is evaluated on WEKA. The classified instances are shown in Figure 6. 25 processes are incorrectly identified with a TP rate of 0.902. The results can be improved with more process instances.

Table 7. Accuracy of the classification model on testing data, source: WEKA

Accuracy of the classification training model	Precision	Recall	F-Score	TP Rate	FP Rate
90%	0.916	0.902	0.899	0.902	0.163

```
=== Confusion Matrix ===

   a   b   <-- classified as
 160   0 |   a = Not Automatable
  25  71 |   b = Automatable
```

Figure 6. Confusion matrix of multi criteria testing model, source: WEKA

5. CONCLUSION

This study set out to develop a multi-criteria model for decision makers in evaluating candidate tasks or routines within processes for automation using RPA. This model is based on 22 criteria to determine their automatability. This research has two main contributions. Firstly, this study synthesizes the understanding of process automation criteria that previously exists in literature. This was done by identifying a group of traits or specifications of different processes and categorized them in to three steps through literature review and survey questionnaire. Secondly, an analysis of the responses from RPA experts results in the creation of a list of process automation criteria. These three sets of criteria are i) process characteristics-based, ii) commercial impact-based, and iii) robotic process mining-based. These three sets of criteria can be used to evaluate business processes in order to identify potential RPA candidates. Criteria such as LP, COG, and DA are proven to be less effective. It has been observed that criteria like CBPS, DUR, and HI are extremely beneficial in RPA outcomes. After coming up with the final set of criteria, the authors created a framework to help future researchers and practitioners. This study contains vital information that professionals may use to prioritize which process to automate, but academics can now identify which areas warrant additional research. Every criterion in this study is centered on RPA projects that are on their early stages of development phase. There are a few more noteworthy limits to this study. First, a literature review was conducted; nevertheless, even with a strict approach, it's possible that some pertinent researches were overlooked. Second, RPA cannot yet be regarded as a mature topic in literature because it is a relatively new technology. As a result, the writers have made do with what has been made accessible thus far. Lastly, it was difficult to locate individuals with a high degree of subject area competence. A couple of suggestions for future work paths are made in light of the limitations. First, the authors are already exploring and evolving this investigation by creating a multi-criteria decision-making model. Further responses from finished RPA projects are anticipated in the future, which will be used to improve and assess the suggested multi criteria model even more. Second, there is still more research to be done on how RPA might be integrated with other technologies. AI and ML to be used with RPA to enhance the cognitive capability of RPA. Finally robotic process mining technique should be in consideration to assess the suitability of RPA projects for automation. It is also intended to assess how well the model and the criteria predict RPA projects in other industrial domains such as finance, auditing and education.

REFERENCES

[1] C. Cewe, D. Koch, and R. Mertens, "Minimal effort requirements engineering for robotic process automation with test driven development and screen recording," in *Business Process Management Workshops*, 2018, pp. 642–648.

[2] W. M. P. V. D. Aalst, M. Bichler, and A. Heinzl, "Robotic process automation," *Business & Information Systems Engineering*, vol. 60, no. 4, pp. 269–272, Aug. 2018, doi: 10.1007/s12599-018-0542-4.

[3] D. Farinha, R. Pereira, and R. Almeida, "A framework to support Robotic process automation," *Journal of Information Technology*, vol. 39, no. 1, pp. 149–166, Mar. 2024, doi: 10.1177/02683962231165066.

[4] I. M. D. Vitharanage, W. Bandara, R. Syed, and D. Toman, "An empirically supported conceptualisation of robotic process automation (RPA) benefits," *ECIS 2020 Research in Progress Papers*, vol. 58, 2020.

[5] F. Santos, R. Pereira, and J. B. Vasconcelos, "Toward robotic process automation implementation: an end-to-end perspective," *Business Process Management Journal*, vol. 26, no. 2, pp. 405–420, Sep. 2019, doi: 10.1108/BPMJ-12-2018-0380.

[6] R. Uskenbayeva, Z. Kalpeyeva, R. Satybaldiyeva, A. Moldagulova, and A. Kassymova, "Applying of RPA in administrative processes of public administration," in *2019 IEEE 21st Conference on Business Informatics (CBI)*, 2019, pp. 9–12, doi: 10.1109/CBI.2019.10089.

[7] M. Lacity and L. Willcocks, "Becoming strategic with intelligent automation," *MIS Quarterly Executive*, vol. 20, no. 2, pp. 1–14, 2021, doi: 10.17705/2msqe.000XX.

[8] P. Hallikainen, R. Bekkhus, and S. Pan, "How OpusCapita used internal RPA capabilities to offer services to clients," *MIS Quarterly Executive*, vol. 17, no. 1, 2018.

[9] S. Hegde, S. Gopalakrishnan, and M. Wade, "Robotics in securities operations," *Journal of Securities Operations & Custody*, vol. 10, no. 1, pp. 29–37, 2018.

[10] P. Hofmann, C. Samp, and N. Urbach, "Robotic process automation," *Electronic Markets*, vol. 30, no. 1, pp. 99–106, Mar. 2020, doi: 10.1007/s12525-019-00365-8.

[11] W. V. D. Aalst, *Process mining: Data science in action*. Berlin, Heidelberg: Springer, 2016.
[12] A. Leshob, A. Bourgouin, and L. Renard, "Towards a process analysis approach to adopt robotic process automation," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, Oct. 2018, pp. 46–53, doi: 10.1109/ICEBE.2018.00018.
[13] J. Wanner, A. Hofmann, M. Fischer, F. Imgrund, C. Janiesch, and J. G.-Klingeberg, "Process selection in RPA projects – towards a quantifiable method of decision making," in *Conference: 40th International Conference on Information Systems (ICIS)*, 2019, pp. 1–17.
[14] H. Leopold, H. van der Aa, and H. A. Reijers, "Identifying candidate tasks for robotic process automation in textual process descriptions," in *Enterprise, Business-Process and Information Systems Modeling*, 2018, pp. 67–81.
[15] K. C. Moffitt, A. M. Rozario, and M. A. Vasarhelyi, "Robotic process automation for auditing," *Journal of Emerging Technologies in Accounting*, vol. 15, no. 1, pp. 1–10, Jul. 2018, doi: 10.2308/jeta-10589.
[16] J. G.-Klingeberg, J. Nakladal, F. Baldauf, and F. Veit, "Process mining and robotic process automation: A perfect match," in *16th International Conference on Business Process Management (BPM)*, 2018, pp. 1–8.
[17] B. Agaton and G. Swedberg, "Evaluating and developing methods to assess business process suitability for robotic process automation - a design research approach," *Master Thesis*, Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden, 2018.
[18] R. Syed *et al.*, "Robotic process automation: Contemporary themes and challenges," *Computers in Industry*, vol. 115, Feb. 2020, doi: 10.1016/j.compind.2019.103162.
[19] A. M. -Rojas, A. J. -Ramírez, J. G. Enríquez, and H. A. Reijers, "A screenshot-based task mining framework for disclosing the drivers behind variable human actions," *Information Systems*, vol. 121, Mar. 2024, doi: 10.1016/j.is.2023.102340.
[20] S. Moreira, H. S. Mamede, and A. Santos, "Process automation using RPA – a literature review," *Procedia Computer Science*, vol. 219, pp. 244–254, 2023, doi: 10.1016/j.procs.2023.01.287.
[21] N. S. S. Elsayed and G. Kassem, "Assessing process suitability for robotic process automation: A process mining approach," *Wirtschaftsinformatik 2022 Proceedings*, vol. 18, pp. 1-16, 2022.
[22] H. Jalonen, "Assessing robotic process automation potential," *M.Sc. Thesis,* Department of Information and Knowledge Management, Tampere University of Technology, Tampere, Finlandia, 2017.
[23] R. Beetz and Y. Riedl, "Robotic process automation: Selection of automatable business processes," *AMCIS 2019 Proceedings*, vol. 4, 2019.
[24] M. Lacity, L. Willcocks, and D. Gozman, "Influencing information systems practice: The action principles approach applied to robotic process and cognitive automation," *Journal of Information Technology*, vol. 36, no. 3, pp. 216–240, 2021, doi: 10.1177/0268396221990778.
[25] K. K. H. Ng, C.-H. Chen, C. K. M. Lee, J. R. Jiao, and Z.-X. Yang, "A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives," *Advanced Engineering Informatics*, vol. 47, 2021, doi: 10.1016/j.aei.2021.101246.
[26] A. Asatiani and E. Penttinen, "Turning robotic process automation into commercial success – Case OpusCapita," *Journal of Information Technology Teaching Cases*, vol. 6, no. 2, pp. 67–74, Nov. 2016, doi: 10.1057/jittc.2016.5.
[27] M. Fakhri *et al.*, "A fundamental review on candidate processes for robotic process automation (RPA) at operational level," *INTI Journal*, vol. 54, 2020.
[28] J. Hindel, L. M. Cabrera, and M. Stierle, "Robotic process automation: Hype or hope?," in *Wirtschaftsinformatik (Zentrale Tracks)*, GITO Verlag, 2020, pp. 1750–1762.
[29] K. V. J. Padmini, G. I. U. S. Perera, H. M. N. D. Bandara, and R. K. O. H. Silva, "A decision support tool to select candidate business processes in robotic process automation (RPA): An empirical study," in *Inventive Computation and Information Technologies*, 2021, pp. 567–582.
[30] V. Leno, A. Polyvyanyy, M. L. Rosa, M. Dumas, and F. M. Maggi, "Action logger: Enabling process mining for robotic process automation," *Demonstration Track of the International Conference on Business Process Management (BPM Demos)*, 2019.
[31] R. Karsi, M. Zaim, and J. El Alami, "Assessing naive bayes and support vector machine performance in sentiment classification on a big data platform," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 990–996, Dec. 2021, doi: 10.11591/ijai.v10.i4.pp990-996.
[32] E. G. Kulkarni and R. B. Kulkarni, "WEKA powerful tool in data mining," *International Journal of Computer Applications*, pp. 10–15, 2016.

BIOGRAPHIES OF AUTHORS

Neelam Yadav received an M.Tech. degree in computer science and engineering from Guru Gobind Singh Indraprastha University of Delhi, India. Her research interests include machine learning, robotic process automation, and artificial intelligence. She can be contacted at email: mail2neelam.sanskriti@gmail.com.

Dr. Supriya P. Panda, M.S., Ph.D. (BGSU, Ohio, USA) is working as a professor (CSE) at Manav Rachna International Institute of Research & Studies at Faridabad (India) since 2016. She started her stint at BITS Pilani, Rajasthan, India wherein she served for ten years. She has numerous research papers in reputed national and international journals to her credit. She can be contacted at email: supriya.fet@mriu.edu.in.