University of Tehran

Iranian Journal of Management Studies (IJMS)

Home Page: https://ijms.ut.ac.ir

Online ISSN: 2345-3745

Analyzing the Research Grant Process in Iran's National Elites Foundation: An Approach Based on Process Mining and Machine Learning

Mansoureh Yari Eili[1] | Jalal Rezaeenour [2*] | Amir Jalaly Bidgoly[3] | Shariar Bijani[4]

1. Department of Computer Engineering, Faculty of Technology and Engineering, University of Qom, Iran. Email: M.yari@stu.qom.ac.ir
2. Corresponding Author, Department of Industrial Engineering, Faculty of Technology and Engineering, University of Qom, Iran. Email: Rezaee@qom.ac.ir
3. Department of Computer Engineering, Faculty of Technology and Engineering, University of Qom, Iran. Email: jalaly@qom.ac.ir
4. Department of Mathematics and Computer Science, Shahed University, Iran. Email: bijani@shahed.ac.ir

ARTICLE INFO

Article type:
Research Article

Article History:
Received 20 September 2022
Revised 11 June 2022
Accepted 21 June 2022
Published Online 28 February 2023

Keywords:
Bottleneck analysis,
decision tree,
Iran's National Elites Foundation,
organizational performance analysis,
process mining.

ABSTRACT

Analyzing the event logs extracted from the process-aware information systems provide critical insights into improving the organizational processes. This case study reports the essential findings and lessons from a process mining project run in analyzing the postdoctoral research grant process in Iran's National Elites Foundation (INEF). Different deductions are reached by exploring the process participants' activities in the INEF web portal, including (1) the organizational inefficiencies exposed through the process mining techniques, where the most time-consuming activities are detected and suggested to the domain experts, and (2) the decision tree technique applied in determining how the successful applicants are scored. The extracted rules indicate an 18% application admission with a final score of more than 403. This article contributes to interpreting the behavioral patterns in INEF and determining who among the applicants has a higher chance of receiving the grant, supporting the policymakers and managers to assign rational budgeting and adopt appropriate human resource strategies.

Cite this article: Yari Eili, M., Rezaeenour , J., Jalaly Bidgoly, A., Bijani , Sh. (2023). Analyzing the Research Grant Process in Iran's National Elites Foundation: An Approach Based on Process Mining and Machine Learning. *Iranian Journal of Management Studies* (*IJMS*), 16 (2), 429-445.
DOI: http//doi.org/ 10.22059/ijms.2022.330862.674766

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/ijms.2022.330862.674766

1. Introduction

Most organizations are equipped with a customized process-aware information system to track business events (Dumas et al., 2018). Capturing and storing all behaviors in these systems leads to a significant volume of execution data, appropriate for analysis from a process viewpoint. Analyzing event logs obtained from these systems provides deep insights into improving organizational processes (Zeng et al., 2013). Process mining (PM) techniques have been applied successfully in different domains, and the resulting insights have enhanced its benefits and capabilities (de Weerdt et al., 2013; Werner, 2017).

The research policy is a prominent factor in the broader political scene. Research in all domains must be considered critical in maintaining and improving growth, welfare, and international competitiveness. This component motivated the growing emphasis on the impact of a research grant and the way this grant can best be attributed to promoting socio-economic progress. This article reports the primary findings and lessons from a PM case study run to analyze the behaviors during the postdoctoral grant process in Iran's National Elites Foundation (INEF).

The INEF is one of the many organizations in Iran responsible for funding research projects and individual researchers. In a quantitative research method carried out by Ruhani et al. (2020), based on the 11 interviews with students with elite status, it was revealed that the research grant process for applicants is quite an ambiguous and complicated task. The essential factors which impact receiving a grant and the steps which an application should take until final status (reject/ accept) are among those which are unknown to many applicants. Many researchers in the humanities disciplines believe that INEF is a foundation merely serving researchers in the engineering disciplines. Another inevitable issue is the fact that with any change in the INEF management, the regulations change as well. This causes dissatisfaction among the applicants because no rational and consistent policy exists. These issues indicate the unknown nature of a research grant process for a group of applicants with the same requests Ruhani et al. (2020).

The administrative structure of INEF is highly bureaucratic due to not having expert staff, which is time and money-consuming (Ruhani et al. 2020). Because of limited financial resources, the pressure on organizations to operate effectively is high. Consequently, they must optimize human resource management and resort to better business processes by having cost reduction in mind to provide high-quality service. Managers in INEF declared that they are interested in identifying organizational bottlenecks (i.e., time-consuming activities/ employees) and activities with the potential for automation or elimination to adopt strategies that fit their objectives. This research, a report of a funded project in INEF, is conducted to determine how the eligible applicants are scored and reveal the feasibility of PM in organizational performance improvement.

The details of systematically collected data related to research grants in INEF are being exposed for the first time, with the objectives of (1) identifying and introducing the influential factors involved in applications' chances of success, (2) discovering the range of different sequences of activity (behavior) regarding the applicants and analyzing the collaboration between departments regarding their cooperative performance. This usually offers beneficial tips to the process owners to improve the service quality and the organizational processes. Consequently, the following research questions are addressed:

RQ-1: What are the influencing factors in receiving INEF grants, and what specific behavior(s) are observed among rejected and accepted cohorts of applicants?

RQ-2: What are the bottlenecks in the organization, and how can they be removed?

These questions would be answered through an exploratory sequence analysis by applying PM and decision tree (DT). To accomplish this, data mining (DM) analysis and PM are carried out using a real-life case study focusing on a sizeable Iranian funding organization's back-office process to broaden the understanding of grants distribution among the eligible applicants.

The rest of this article is organized as follows. The literature is reviewed in Section 2, and the case study is introduced in Section 3. Section 4 describes the event data collection and event log preparation, while Section 5 presents the findings from the above-mentioned questions of interest. Section 6 discusses a few suggestions to process improvement for the process owners and PM experts. Finally, Section 7 concludes the paper.

2. Literature Review

This work extends previous studies on research grant analysis, educational PM, and financial PM.

2. 1 Research Grant Analysis (RGA)

Some studies have assessed the research grant's impact on the quantity and quality of the peer-reviewed articles written by the funded scientists. The detailed information on 20,476 research grants by the Swiss National Science Foundation (SNSF) submitted between 2005 and 2019 were used by Heyard and Hottenrott (2021) to estimate the grant award impact on the count of articles, citations, and relative citation ratios. The negative binomial and linear mixed models' outcome revealed that the funding program facilitates 1.21 times higher annual publications. In this context, a moderate researcher publishes 4.64 articles on average without SNSF funding, while this number for a funded researcher is 5.6.

Linear regression analysis and t-test were adopted by Berg and Ashurst (2019) to assess the NIH funding trends in general surgery from 2008 to 2017, and the outcome revealed a gender and degree disparity: 76.33% of the awards assigned to males and 58.33% to those holding a medical doctor degree.

The influential factors in research quality were assessed by Jung et al. (2017), where the data from scientists funded by the National Research Foundation (NRF) of Korea from 2003 to 2009 were used. To estimate the NPF's impact, the academic disciplines and the obtained Impact Factors (Ifs) were analyzed through Poisson, ordinal logistic, and logistic regression methods. Their findings indicated that faculty members with a high count of graduate students, with many coauthored articles, and funded researchers have the potential to be published in renowned journals.

The relation between the funded volume provided to 12,720 researchers in Québec and their scientific output was assessed by Mongeon et al. (2016). The local regression and the Cobb-Douglas production function were employed to determine the upward, downward, and constant marginal returns on annual research funding. The article count, the average relative citations, and the top-cited articles' count revealed a non-proportional upward trend in funding, up to a specific breakpoint after which the downward trend began.

Wang and Shapira (2015) analyzed the funding of 89000 nanotechnology articles in one year to assess the correlation between funding and publication impacts. They draw on a two-stage regression model to test the two bibliometric indicators: journal IF and funded articles' citation count. As to journal ranking and citation counts, they found that grant-sponsored researchers are published with high-impact publications.

The findings of the mentioned studies correspond to the Anesthesia education and research Grant Program (Pagel & Hudetz, 2015), the Danish Council for Independent Research (Bloch et al., 2014), the Mycology research investments in the UK institutions (Head et al., 2014), the Chilean FONDECYT (Benavente et al., 2007), and the natural sciences and engineering research council of Canada (Godin, 2003), which reveal that the research grant programs have their impact on the scientific publication rate and increase citation count. This issue indicates that a grant's impact goes beyond the quantity and promotes circulation and quality.

2.2 Educational Process Mining (EPM)

PM has been applied to (1) discover learning traces by detecting learning trends and student profiles, (2) monitor student's behavior by diagnosing the bottlenecks, analyzing performance, and checking conformance, and (3) improve learning processes by recommending an appropriate trace (Bogarin et al., 2018). The focus of the EPM studies is on different educational environments (e.g., e-learning, undergraduate courses, and universities) rather than the research grant programs.

Sonnenberg and Bannert (2019) used a conformance checking approach to measuring the metacognitive prompting effect on the learning process of 69 university students in two hypermedia learning sessions, educational psychology being the topic. The findings revealed the benefits of assessing instructional support where the sequential structure of learning processes is of concern. Van den Beemt et al. (2018) explored the relationship between learning behavior and learning progress in the MOOC to gain insight into the activities of passing and failing students. The cluster analysis and PM were combined for behavioral detection in four clusters of students, and the results suggested teacher guidance. Douzali and Darabi (2016) employed the PM in parallel to data mining techniques on the Honors Program admission process at the University of Illinois, Chicago. They assessed the educational history to identify the students with the highest potential for admission. SoftLearn, the only PM tool for the educational domain, developed by Barreiros et al. (2014), is based on the genetic miner and social network. This tool allows teachers to visualize the students' learning traces and enhance the overall learning process.

2.3 Financial Process Mining (FPM)

The successful application of PM in different domains is well known through many studies, but a recent literature review by dos Santos Garcia et al. (2019) revealed that only 6.5% of the studies have focused on the PM techniques in finance institutions. The FPM studies are carried out on insurance claim handling, automated teller machines, banking contact centers, loan approval, etc. (de Leoni et al., 2016; Lakshmanan et al., 2015; Mahmood & Shaikh, 2013; Peters et al., 2013). To the best knowledge of the researchers here, to date, the FPM has not been applied to the research grant process.

Werner (2017) considered data dependencies related to the accounting structure of events to discover auditing processes in fraud detection. The generated models provide accurate control-flow information with a lower complexity vs. those generated by timestamp dependencies. Conforti et al. (2015) proposed a recommendation system for risk minimization by running multiple process instances. They sought to support people involved in the claim handling process of a major insurance company in making risk-informed decisions. For each running case, the risk prediction, in terms of error probability if the case is executed in such a way, is calculated through the DT on top of a YAWL plug-in named Map Visualizer.

The studies examining the factors that influence research productivity are numerous, but they do not consider the factors that influence the optimization of allocating research grant process. This study contributes to the available literature by assessing the primary rules of receiving a research grant and determining the business process improvement opportunities in reducing the time and cost of process execution through process mining techniques. In this context, the novel data set from INEF allows us to have a more comprehensive assessment of the determinants of qualified applicants, where different elements related to the individual researcher are involved. A comparison between the mentioned studies above is tabulated in Table 1.

Table 1. Comparison of Different Studies

Authors/ year	Field of study	Technique	Factors analysis	Behavior analysis	Bottleneck analysis	Institution	Explanations
Heyard & Hottenrott (2021)	RGA	Longitudinal regression models	-	-	-	Swiss National Science Foundation	Impact of grant on the productivity (number) and quality (citation) of research
Berg & Ashurts (2019)	RGA	Linear regression analysis	+	-	-	National Institution of Health in US	Statistical analysis on the degree and gender disparity that exist in the total grants
Jung et al. (2017)	RGA	Poisson model, ordinal logistic, regression	−	-	-	Mid-Career researcher program	Impact of grant on the quality of research (journal IF)
Pagel & Hudetz (2015)	RGA	Kolmogorov-Smirnov, Mann-Whitney, Kruskal-Wallis tests	−	-	-	Foundation for Anesthesia Education and Research (FAER)	Impact of grant on the productivity (count) and quality (citation and h-index) of research
Wang & Shapira (2015)	RGA	Regression model	-	-	-	Nanotechnology researchers	Impact of grant on the publication impacts (funding acknowledgments in articles)
Bloch et al. (2014)	RGA	Mixed methods	-	+	-	Danish council for independent Research	Impact of grant on productivity and quality of research between accepted and rejected applicants

Table 1. Comparison of Different Studies (Continoud)

Authors/ year	Field of study	Technique	Factors analysis	Behavior analysis	Bottleneck analysis	Institution	Explanations
Mongeon et al. (2016)	RGA	Cobb-Douglas production function and local regression	–	-	-	Québec funds over a 15-year period	Impact of grant on the publication (funding acknowledgments in articles)
Head et al. (2014)	RGA	Statistical analysis	-	-	-	Mycology research investments in the UK institutions	A systematic analysis of grants by the research field and institution
Benavente et al. (2007)	RGA	Regression Discontinuity design	-	-	-	Chilean National Science and Technology Research Fund	Impact of grant on the publication
Godin (2003)	RGA	Statistical analysis	-	-	-	The Natural Sciences and Engineering Research Council	Impact of grant on the publication
Sonnenberg & Bannert (2019)	EPM	Process mining	-	+	-	learning process of 69 university students	a conformance checking approach to measuring the metacognitive prompting effect on the learning process
Van den Beemt et al. (2018)	EPM	Process mining	-	+	+	MOOC	Analyzing learning behavior to investigate the activities of passing and failing students
Douzali & Darabi (2016)	EPM & RGA	PM and DM	+	-	-	Honors Program (HP) in the University of Illinois	The selection process discovery for admission into HP
Barreiros et al. (2014)	EPM	PM	-	-	-	virtual learning environments	SoftLearn, only PM tool for educational domain
Werner (2017)	FPM	PM	-	-	+	Banks	Financial auditing processes to fraud detection
Conforti et al. (2015)	FPM	PM	-	-	+	Claim handling process	A process-aware recommendation in insurance company
Present study	RGA ,EPM ,FPM	PM and DM	+	+	+	INEF	Investigation of: -significant factors on receiving research grant -process improvement opportunities

3. Introducing the Case Study

Iran's public and private entities provide financial grants for research, science, and technology development. The INEF, one of the statewide non-governmental funding bodies in Iran, was founded in 2006 to provide opportunities for talented students, faculty members, researchers, and innovators to enhance their research and training skills. In 2020, it funded more than 5000 research opportunities and channeled the related grants in a variety of manners such as salary, gratuitous loans, and supplying laboratory facilities. With a sizable volume of data related to thousands of applicants stored in the INEF database, the organization had no clear idea of the details of research grant processes and how accurate performance estimations for different request types were made. The consultations with the INEF managers and stakeholders revealed that the processes were not fault-free (e.g., the existence of complex bureaucracy that increased process time and cost due to unnecessary activities). Because the count of grants that this foundation is to allocate is not inexhaustible, modifying all the practical issues will reduce the response and process execution time/cost through optimal human resources. All these will give the applicants a better idea of how a research grant is allocated.

Access to the data requested for this study was restricted to the Chamran Prize process, a well-structured process of postdoctoral research grant where data is collected more comprehensively. The

Chamran Prize, the *post-doc process* in this article, provides an annual salary for PhD graduates seeking to follow their postdoctoral research careers.

The first step in this process is the *submit request*, where applicants should upload their documents, including educational background, through the Sina portal. Each request is assigned a unified identification code. In the second step, the *automatic scoring*, the grant application is scored by the scoring mechanism described in detail on the INEF website. The candidates must get the minimum score. If the score is higher than the threshold, the request is directed to further steps; otherwise, a *notification about rejection* will be sent to the applicant's profile.

A general request trace is flowcharted in Figure 1: the request goes through *check by the provincial branch* and *working group reviewing* in different departments for more review. Then the applicant may be rejected or allowed to continue the procedure until acceptance. These activities are often followed by *user information modification*, requesting a proposed change to the documents. The explanation of activity labels is expressed in Appendix 1.

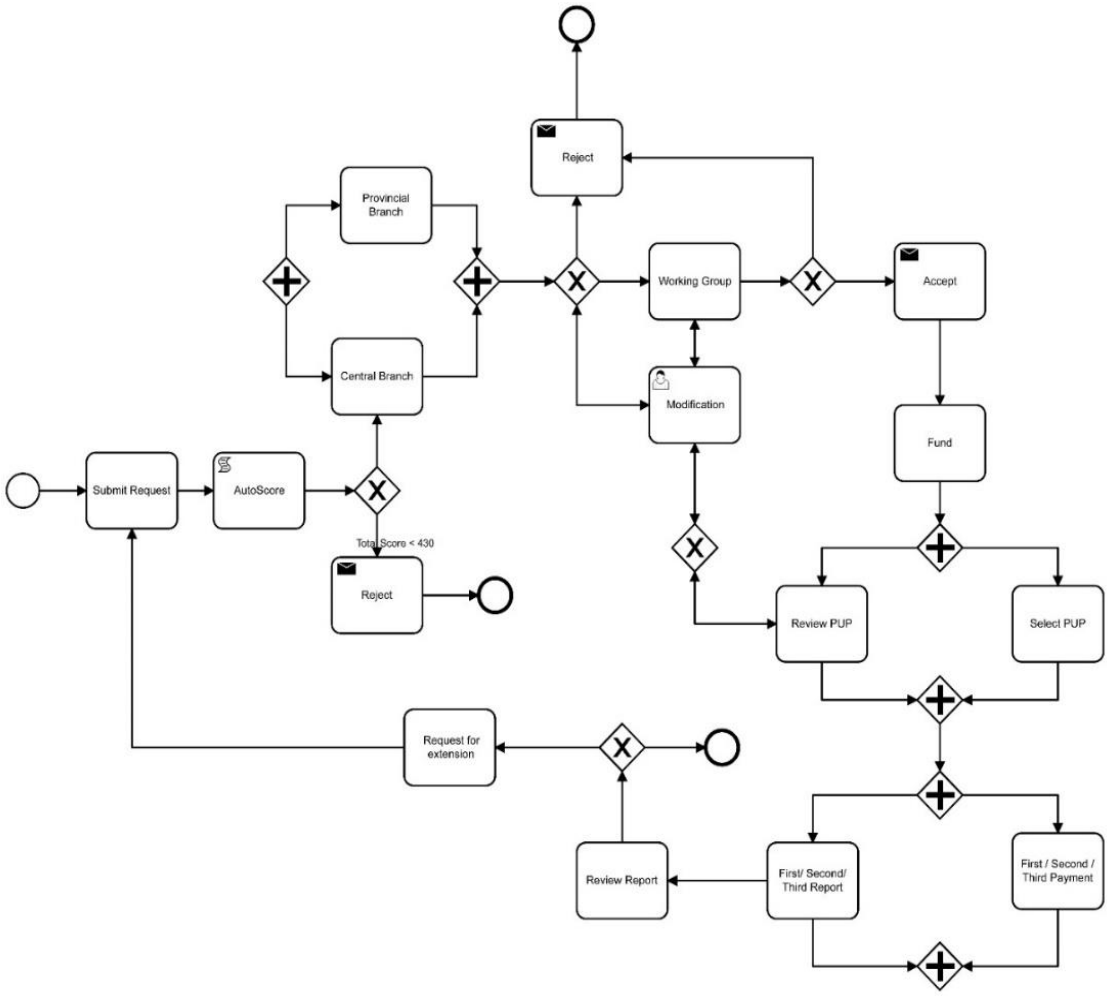

Figure 1. The General Procedure of the Post-Doc Research Grant Process

In this case study, the PM2 process mining methodology introduced by Eck et al. (2015) is adapted consisting of (1) setting up the research questions, (2) identifying and extracting the relevant data, (3) building event logs from the raw data, (4) applying process mining techniques into the event logs and gaining insights on process performance and compliance, (5) analyzing the findings to improve ideas that would be contributive to the project's objectives, and (6) detecting problematic running cases, resources, and activities. The overview of this study design is shown in Figure 2.

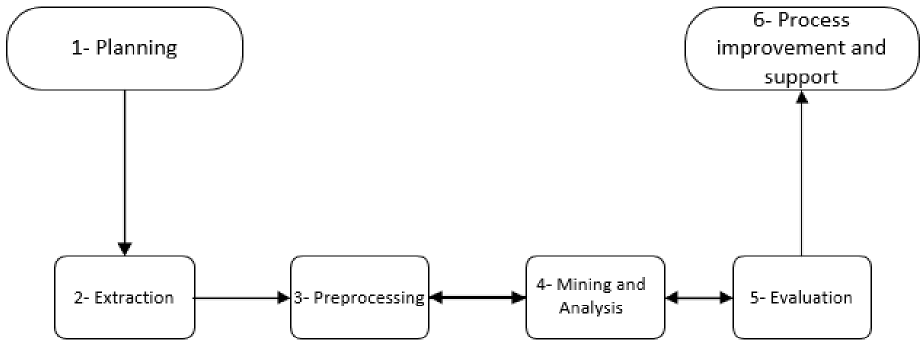

Figure 2. Overview of the Methodology Applied in This Study.
(The process mining technique is applied to the INEF data. The resulting data were transferred to the event log to provide input for analysis.)

4. Data Preprocessing

One of the first phases in any PM project is the data preprocessing and event log building, which is the input of any PM techniques. The INEF database keeps track of every step during the request life cycle by recording the actions, the resources, and the activities' occurrence time. These data constitute the cornerstone of the PM techniques. The data are provided in a tabular format (Microsoft Excel), where the applicant's information table, categorized by user ID, varies from the event log table, categorized by case ID. Each applicant with a unique user ID can submit different requests identified by case IDs. The lack of data integration between these tables is one of the significant challenges in data preprocessing.

In the first step of this phase, the data were extracted and converted into a standard event log format (XES). To enhance the performance analysis considered in the control-flow perspective, both the start and the end timestamps should be involved (de Weerdt et al., 2013). Many noise factors might be present in the dataset (e.g., incomplete, incorrect, and duplicate data entry) (Suriadi et al., 2017). This case study dealt with the multiple recording noise, where two or more records refer to the same event due to technical reasons.

All the requests involved in all process steps, from submission to notification regarding rejection/acceptance, were considered. The applicants who had recently submitted requests (386 running cases) or had incomplete requests for more than 6 months (1059 incomplete cases), that is 62% (1445 out of 2351 requests), were withheld. With these issues at hand, the final event log consisted of 906 execution traces and a total of 12148 events of all cases recorded from July 21, 2018, to July 19, 2021.

The objective was to compare the accepted applicants' behaviors vs. rejected applicants. Therefore, the event log was segmented into three separate cohorts for analysis based on whether the applicants' requests were accepted; if so, they canceled the request for not going through further steps. This approach is common in performing process behavioral comparison, known as process variant analysis in PM (Van den Beemt et al., 2018). The segmentation above provides relevant insights concerning the research questions. A small portion of the post-doc event log is given in Table 2, and the statistical details are provided in Table 3.

For data cleaning purposes, the outlier values in the applicants' information, outside the expected range (e.g., GPA>>20), or feature values recorded only for a low count of applicants (e.g., Science Olympiad) were deleted. Based on the ISC university ranking system[1], the university information of applicants was classified as $C_1, C_2, \dots, C_6$. According to the GPA scale, the GPA score of 10, …, 20 was replaced with $A, B, \dots, F$. Because the *Label* attribute was the *status of request*, all users with an *unknown* status were omitted in rule discovery, which improved the DT interpretability.

1. https://www.msrt.ir/en , https://wur.isc.ac/Home/WorldIslamicUnivRankings?year=2020

Table 2. An Example of INEF Event Log

Case ID	Activity	Timestamp-start	Timestamp-end	Resource
7364994	Submit	9/16/2020 5:13:27 PM	9/16/2020 5:13:27 PM	Applicant
7364994	Auto score	9/16/2020 5:13:27 PM	9/16/2020 5:13:27 PM	System
7364994	Central branch	9/16/2020 5:13:27 PM	9/19/2020 4:14:10 AM	Staff1
7364994	Provincial branch	9/19/2020 4:14	9/19/2020 7:55	Staff2
7364994	Central branch	9/19/2020 7:55	9/20/2020 8:21	Staff1
7364994	Modification	9/20/2020 8:21	9/24/2020 8:15	Applicant
7364994	Working group	9/30/2020 6:22	10/26/2020 8:26	Staff3
7364994	Reject	12/30/2020 10:26	12/30/2020 10:26	Staff3

The personal data available in the dataset made it possible to realize the applicants' population. Of the 906 requests submitted to the INEF, 60% were male (women were founded less), of which 19% had a GPA of A and 48% a GPA of B. As to their institutions, 43% graduated from first-ranked (C1) and less than 20% from C4-C6 universities. The majority of grants were given to Chemistry (17%), Electronic Engineering (9%), and Physics/ Mechanical engineering (8%) research fields. The humanities fields received the least grants (less than 10% in aggregate). According to the Final Score (FS), less than 5% of grant applications scored FS> 550, 18% scored a final between 450 and 550, while 65% scored between 250 to 450. Finally, about 12% of applications scored FS<250. The funded applicants made 18%, the rejected 70%, and the canceled 12%. These are illustrated in Figure 3.

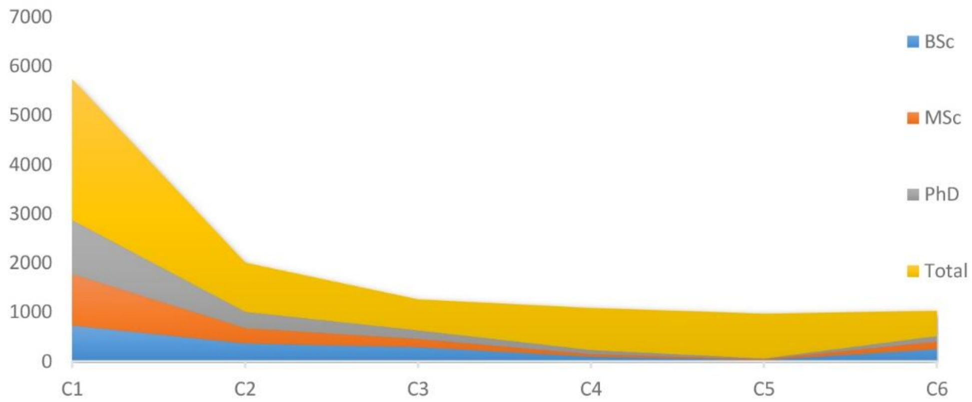

Figure 3. Distribution of Applicants Based on the Grade and University Ranking

5. Case study Findings

This section explains how each research question was addressed, and the results were obtained. The Disco[1] was selected for PM analysis because instead of finding choices and parallelisms, the process map discovery and performance insights were the priorities of this study. Disco is primarily based on an advanced version of the Fuzzy Miner from Günther and Van der Aalst (2017), which contributed to discovering process execution sequences and loops, determining waiting times, bottlenecks, and potential causes of delay by highlighting frequent activities.

Fuzzy miner deals well with loop and noise in less structured behaviors, where the event logs cannot easily be summarized in a structured process model. Omitting the unimportant edges and clustering the highly correlated nodes makes process model interpretation easy. However, there exist some limitations to this miner in the practical application. The extracted model cannot be converted to Petri net, limiting a comparative evaluation to other miners (de Weerdt et al., 2013). The discovered fuzzy model does not contain semantics and only applies to indicating the correlations among activities. The fuzzy model cannot analyze the non-free choice behaviors and lacks efficiency in running time compared with other algorithms like the Heuristic Miner (Ajayi et al., 2019). The RapidMiner Studio[2] was applied for DM and rule discovery.

1. https://fluxicon.com/disco/
2. https://rapidminer.com/

5.1 RQ-1 What Specific Behavior(s) and Case Attribute(s) are Observed Among a Different Cohort of Applicants?

In the exploratory phase, 321 process variants among the 906 cases were observed in the event log. To address RQ-1, an endpoint analysis was carried out to segment the cases into three cohorts:

- Cohort 1: Rejected application — (70% of log)
- Cohort 2: Accepted application — (18% of log)
- Cohort 3: Cancelled application — (12% of log)

This segmentation of cases was to compare behaviors with varying levels of steps taken to pass the one-year grant finally.

Cohort 1 consisted of the cases where their requests were not approved because of receiving a low reviewing score. Rejection was made through two different traces: (1) a few applications (33 cases) were removed in the first stages (i.e., *AutoScore* or *Reviewing by Provincial Branch),* and (2) the remaining (635 cases) were rejected after being *reviewed by the working group*, indicating that most applications went for further review. The reviewers in the working group score each application by considering the predefined criteria (e.g., educational and research background). The discovered process model for this cohort is shown in Figure (4-b). The numbers in the model are the cases following a specific trace through the model. Cohort 1 consists of 89 trace variants with the median and means case duration of 88.1 days and 14.6 weeks, respectively.

Records in Cohort 2 were characterized by taking at least one year to complete the post-doc course, which is an inherently long duration compared to other cohorts (Table 3). The process model of this cohort was split into two phases: (1) acceptance and (2) taking a one-year post-doc grant (Figure 4-a). The first phase is similar to the second trace of Cohort 1, where the application goes for further review by the working group. In the end, *notification about acceptance* will be sent to the applicant. Upon acceptance, the applicant should select a host professor and submit a formal grant proposal. Following the research grant proposal evaluation (*PUP* activity), the applicants should provide the research reports on meaningful outcomes quarterly. Cohort 2 consisted of 119 trace variants with the median and mean case duration of 37 weeks and 12 months, respectively.

Cohort 3 consisted of the cases where the applicant withdrew the request even after acceptance. The automatically discovered process model is shown in Figure 4-c. Of 105 canceled items, most (94 applicants) were canceled in the initial reviewing phase and 11 after acceptance. According to the process owner, these applicants might have found a job or joined a university as faculty members.

Compared with other cohorts, the volume of process variants in Cohort 2 was considerable. The core of the different variants was formed from the loops caused by *modifications by the user*. Execution of this activity in a random irregular manner led to a new process variant. The miner considered two traces $T_1 = \langle A, B, C, F, C, D, F, D, F, E, F, N \rangle$ and $T_2 = \langle A, B, C, D, F, D, E, N \rangle$ as different variants, while the general procedure was similar and completed by *acceptance*. After acceptance, *modification* could occur following the *SelectPUP* and *Review report* activities. This execution led to correct irregular patterns, known as *exceptional traces* (Cheng & Kumar, 2015). The exceptional traces were considered in this study due to their frequent sub-sequences.

Table 3. The Event Log and the Cohort Summary

Attribute	**Log**	**Cohort 1 (rejected)**	**Cohort2 (accepted)**	**Cohort 3 (canceled)**
Number of cases	906	668	133	105
Number of events	12148	7771	3499	882
Duration of cases	50 W*	14 W*	12 M**	41 W*
Activities	69	19	61	28
Event per case (max, min, mean)	64, 11, 3	41, 10 ,3	64, 26, 8	18, 9, 7
Number of trace variants	321	89	119	11
Start and end time	July 21, 2018- July 19, 2021			

*W: week, **M: month

5.1.1 User Attributes Analysis (Rule Discovery)

The DT learning approach was adopted to discover the influential rules promoting the chances of receiving a grant application in INEF. For this purpose, the three final statuses of applications (rejected, accepted, and canceled) were set as the label variable. The response variables consisted of GPA, university ranking, research/educational background score, the selected students (yes/no), past research projects in INEF (yes/no), and the final score attributes.

DT approach requires the following parameters to set. *The gini-index* criterion is the attribute selection score measure that facilitates the more significant distributions (it favors larger partitions). *The final score* attribute is the root node for splitting tree, the confidence set to 0.1, a minimum leaf size equal to 2, and the minimal gain parameter equal to 0.001, to obtain a DT with a maximum depth of 10 (see Figure 5).

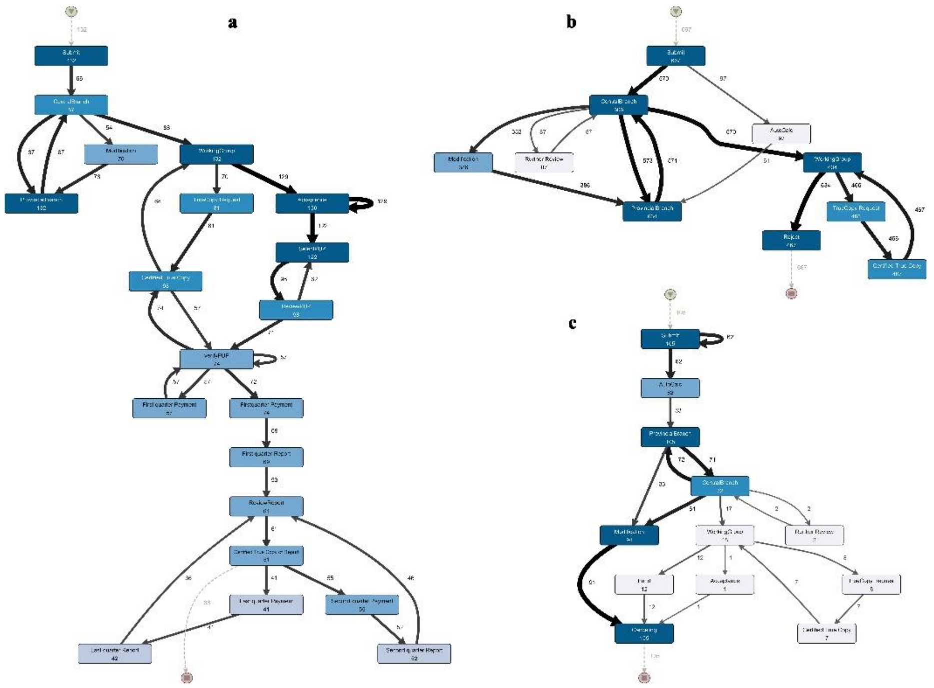

Figure 4. The Discovered Process Model for A: Accepted, B: Rejected, C: Canceled Application

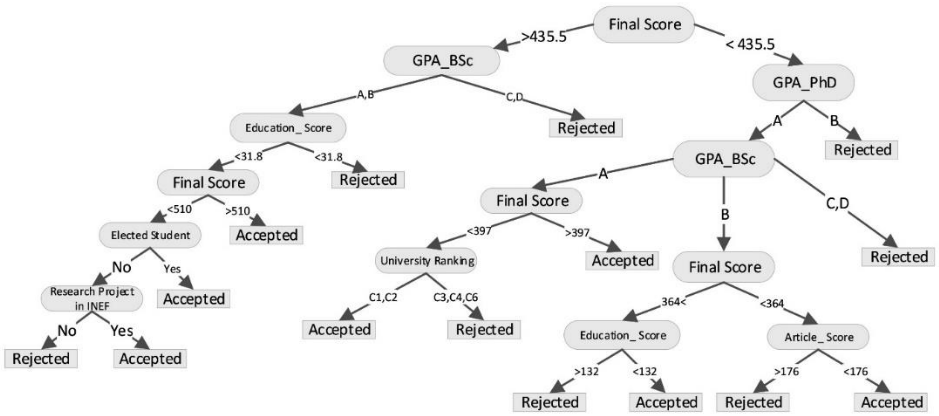

Figure 5. The Summarized DT with the Gain Parameter Set to 0.001

The obtained DT was very detailed and possibly complicated to be represented. For a better understanding of the DT structure and further application in the recommendation, different rules were extracted by the *tree to rules* operator in RapidMiner. Among the total of 80 extracted rules, the most informative high confidence rules consisted of:

Rule 1: *if* Final Score > 403.5
and GPA in BSc = B *and* University ranking in BSc = C1
or Educational Score < 104 *and* Elected Student = Yes, *then* accepted.
Rule 2: *if* Final Score <= 403.5
and GPA in PhD= A *and* previous research collaboration with the INEF= Yes, *then* accepted
Rule 3: *if* Final Score < 397, *and* University ranking in BSc = C3, *and* GPA in MSc= B, *and* Writing/ Translating a scientific book= Yes, *then* accepted
Rule 4: *if* 165 <Final Score < 397, *and* GPA in BSc= C, *and* Educational Score < -25 *and* Article Score > 442, *then* accepted
Rule 5: if Final Score > 403.5, and Article Score > 436, *and* GPA in PhD = A, *and* Writing/ Translating a scientific book= Yes,
or previous research collaboration with the INEF= Yes, *then* canceled.
Rule 6: if 367<Final Score < 403, and GPA in PhD= A, and Writing/ Translating a scientific book= No, then rejected.

According to these six rules, the final score of the most funded applications is more than 403. When the educational score is low, the *elected student*, the *previous research collaboration with INEF*, or *writing a scientific book* lead to acceptance. Based on Rule 4, a high *research score*, estimated based on the count of publications, is highly contributive in funded applications with a low educational score. The most common condition among rejected applications is a low *educational score* (e.g., low university ranking, GPA, etc.). Despite the acceptable GPA or university ranking of the applicants, a lack of research background leads to application rejection (Rule 6). In the branches with a cancelation end node, the applicant withdraws their request, despite obtaining a final score higher than 4.3 (i.e., the conditions by a high probability of acceptance).

5.2 RQ-2 What are the Bottlenecks in the Organization, and How Can They Be Improved?

5.2.1 Bottleneck Analysis

In organizations like INEF, the essential performance indicator is the rapidity in response to the requests, measurable by the activity duration through performance analysis. The time interval estimates the activity duration from receiving to handling a request in each process instance. By analyzing data during the 07.21.2018 -10.18.2020, it was determined that (1) *reviewing by the central branch*, (2) *reviewing by the working group*, and (3) *document modification by a user* were among the most time-consuming tasks.

As observed in Figure 4, before assigning the application request to the *provincial branch*, it is checked by the *central branch*. Even after the provincial branch checks, the central branch takes the application back to decide whether to reject or assign it to the *working group*. Staff and process owners were aware and dissatisfied with this volume of task handover. As observed in the organizational mining results (Figure 6), the reiterated handover between branches is evident (i.e., the arrows in the social network correspond to the undertaken activities by the resources). Specifically, the *Central Branch* has the most intensive arrows in the social network, followed by the *Working group* and the *Branch2*. By nullifying the intermediary rule of the *Central Branch*, the case handling duration of approximately 1/3 of the applications reduces from 8 to 2 days. In the modified process, the submitting request is directly assigned to the *provincial branch*.

The Working Group is responsible for handling activities like *Reviewing by a working group* (36 days mean duration) and *Reviewing PUP* (41 days mean duration). According to process owners, the irregularity and disorder in working group meeting dates lead to a high waiting time in the application review. The count of submitted requests is expressed in Figure 7-A, and the count of working group meetings per month in the three-year study is expressed in Figure 7-B. As observed, during March, June, August, and September, the organization encounters large volumes of requests, thus, holding more meetings. Due to no meetings, applications registered in October and November would be reviewed in December. According to Figure 6, the network efficiency in the research grant process corresponds with that of the working group. An organization should hold monthly meetings to improve the response process to be effective.

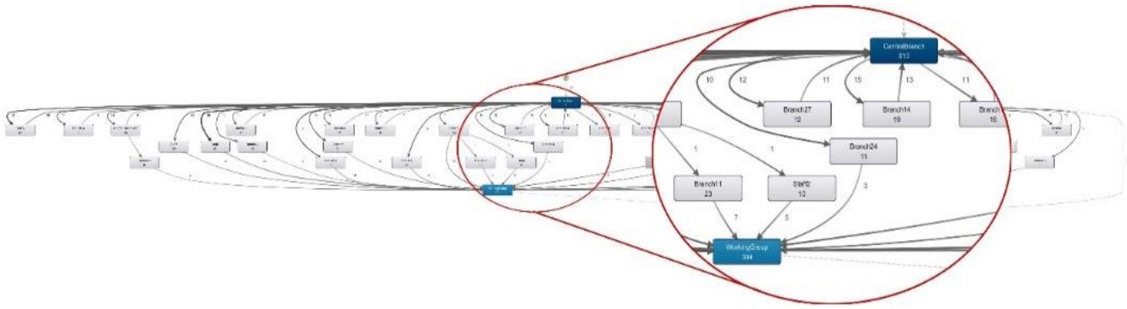

Figure 6. The Organizational Perspective of the Post-Doc Process in the INEF

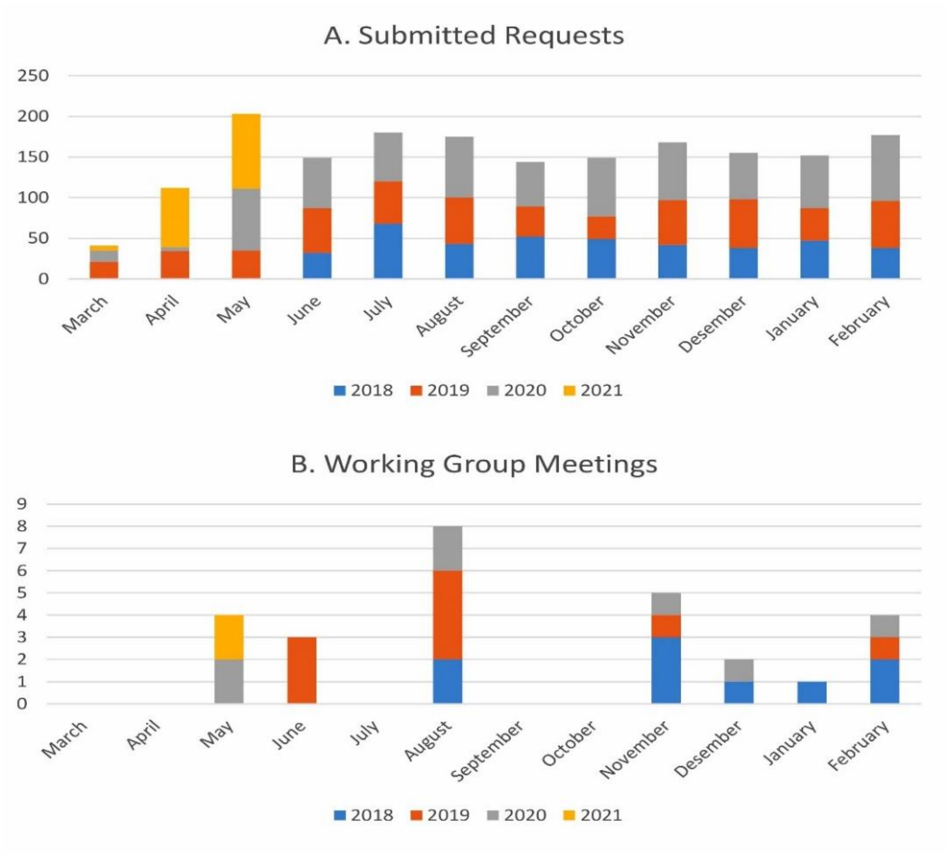

Figure 7. A) The Count of Submitted Requests, B) The Count of Working Group Meetings per Month

6. Discussion

The key lessons learned based on the data-driven insights obtained in this PM case study run on a research funding body and the feedback from process owners throughout the project are synthesized in this section. By contrasting the actual behaviors registered in the event log, expectations and requirements of different guidelines and insight become evident. According to the organization's management, the results of this case study are outstanding.

6.1 Creation of Current State Processes and Identifying the Areas of Improvement

Organizations focus on the improved *to be* process, with less interest in assessing *as is*. Realizing the process at hand is contributive to knowing whether to invest in improvements, where performance problem and opportunities are inevitable, and how would it affect the process throughout the organization. Consequently, some enterprises either resort to the current process analysis and adopt shortcuts therein, or hire consultants at high cost to analyze the *as is* process. PM can assist organizations by extracting data from information systems and provide a visually appealing data-driven scene on how processes are executed.

Based on the endpoint activity, the event log is segmented into three cohorts: Rejected (70%), Accepted (18%), and Cancelled (12%) requests to comprise behavior with varying levels of steps taken to receive a one-year grant finally. By considering the performance perspective and sharing the description of how a business process is being executed with the process owners, it is possible to identify and eliminate unnecessary points for which the highest efforts are made. After eliminating the intermediary rule of a *central branch* within the 10/18/2020 to 07/19/2021 study period, the count of activities and the meantime an application reviewed in the system decreased to 42 activities (from 51 in the last version) and 17 working days (from 68 days in the last version), respectively. In the modified version, from 709 cases, 530 were responded in 2 days, while in the last version, from 1553 cases, 647 were responded in more than 8 days. The disclosure of multiple inefficiencies provides a good opportunity for management to become aware of employees' handover process and improve the organizational process by enhanced instructions.

6.2 Identifying Target Communities to Expand Funding Opportunities

Assessing the traces passed by each cohort next to the contextual information analysis of applicants contributes to identifying the target community and adapting a grant allocating strategy to the eligible applicants by considering the limitation in financial resources that is innate in any organization. It is revealed that the graduates from first-ranked universities and leading Iranian institutions for science and research receive over two-thirds of the total grants. In contrast, lower-ranked institutions often receive only a small portion of the grant. Decentralization and supporting researchers in areas of high need (underprivileged states) or neglected research disciplines (e.g., social science, management, etc.) are matters of focus for policymakers in the short- and mid-term.

6.3 Prescriptive Process Models Should be Adopted to Generate Process Improvement Recommendations Automatically

PM techniques are descriptive in nature (Yari Eili & Rezaeenour, 2021) because they provide insights into the available occurrences. These techniques typically do not automatically recommend actions to be taken for process improvement. By resorting to the PM results, the data analysts introduce improvement opportunities. PM is making advances in identifying and improving opportunities automatically. The PM-based recommender system is an intriguing and new topic, which will be assessed in our upcoming study.

The limitations here are multifold. First, because of privacy preserving issues, direct access to the database was not allowed. The IT department manages the database in INEF that provides access to the data for research. A non-disclosure agreement is concluded in this case study to prevent raw data publication, through which the release of applicants' data is protected. The required data for process mining project was first explained to the IT department staff in detail, and then they provided us the required data scattered through different tables. The authors had to apply the low-quality raw data extracted by the IT department staff, which had to go through the preprocessing phase. According to the *Data pre-processing* subsection, the applicant's information table is categorized by user ID, and the event log table is categorized by case ID. Each applicant with a unique user ID can submit different requests identified by case IDs. Linking the diversity in content, format, and structure of data sources that are not recorded in a process-oriented manner was a challenging task in this research.

Second, despite the advances made in process mining techniques, efficiency in event log extraction phase is low and has had no reasonable progress over the years. Researchers usually encounter event logs that are not stored in the proper format; hence data transformation is one of the most time-intensive stages in PM projects. Efforts should be made to extract the correct data from INEF in an event log format. As commonly stated, 80% of time and 50% of the costs are attributed to data extraction and preparation phases. This project required manual effort in an ad-hoc manner, where many iterations are necessary to assure the appropriate receipt of the data in their optimal format. These and similar limitations lead to issues concerning data quality, thus an effect on the accuracy of process mining results. The lack of a comprehensive tool for event log building is critical in process mining projects.

Third, the lack of access to the financial information database system in INEF is another limitation. This study was carried out in the Sina portal, where the financial sector data and their contributions to areas like the process performance are unknown. This study does not concern some of the financial process-related data, including activities an application should assigned in the sub-process of *Fund* activity. Further research should focus on other sub-processes in INEF to clarify the duties of participants involved in the research grant process. In this endeavor, further access should be allowed to other INEF databases.

7. Conclusion

This paper presents the first comprehensive assessment of the identified strengths and weaknesses through data analysis on the research grant process in INEF. The objective was to discover the process model, position bottlenecks, and improve the processes. In this context, the core task was to characterize the research grant process within the PM results.

The discovered process model of frequent behaviors describes the order of performed activities and to which results from these sequence of activities lead over time. This description can expose some organizational weaknesses and priority areas for policymakers to improve organizational process (re)design which would gain more planned and targeted student support. Consequently, the findings of this study could be applied as a beneficial resource to improve the user experiences of those who plan to submit applications to INEF. It is found that for students who graduated from lower-ranked universities, certain factors influence the application score significantly. By addressing the identified common factors, applicants will be encouraged to improve their grant applications and the possibility of being funded.

Because this is the first national systematic analysis in this context, comparing studies in other countries is possible. First, the focus of this study was on the post-doc data, one of the several types of grants awarded by INEF. Additional high-quality process data about other research grants would contribute to a better understanding of the accepted applicants' behaviors and features. Second, further research can focus on the early detection of cohorts of applicants based on their contextual information, identifying the potential applicants, predicting the output of the running cases, and providing suggestions that would support mid-term or long-term planning and budgeting. Finally, future researchers can apply the discovered rules and behaviors to conduct a process-aware recommender system based on DT.

Acknowledgment

This research was supported by the fund of Iran National Elites Foundation (INEF) under the research project 'Decision support systems: Organizational PM with experimental implementation in the INEF.'

References

Ajayi, L. K., Azeta, A. A., Owolabi, I. T., Damilola, O. O., Chidozie, F., Azeta, A. E., & Amosu, O. (2019). Current trends in workflow mining. *Journal of Physics: Conference Series*, *1299*(1), 012036.

Barreiros B. V., Lama, M., Mucientes, M., & Vidal, J. C. (2014). Softlearn: A process mining platform for the discovery of learning paths. In *14th international conference on advanced learning technologies* (pp. 373–375). IEEE, Athens, Greece.

Benavente, J. M., Crespi, G., & Maffioli, A. (2007). *The impact of national research funds: An evaluation of the Chilean FONDECYT*. Office of Evaluation and Oversight. Inter-American Development Bank. Available at: https://publications.iadb.org/en/impact-national-research-funds-evaluation-chilean-fondecyt

Berg, E. J., & Ashurst, J. (2019). Patterns of recent National Institutes of Health (NIH) funding in general surgery: Analysis using the NIH RePORTER system. *Cureus*, *11*(6).

Bloch, C., Sørensen, M. P., Graversen, E. K., Schneider, J. W., Schmidt, E. K., Aagaard, K., & Mejlgaard, N. (2014). Developing a methodology to assess the impact of research grant funding: A mixed methods approach. *Evaluation and Program Planning*, *43*, 105-117.

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1), e1230.

Cheng, H. J., & Kumar, A. (2015). Process mining on noisy logs—Can log sanitization help to improve performance? *Decision Support Systems*, *79*, 138-149.

Conforti, R., de Leoni, M., La Rosa, M., Van der Aalst, W. M., & ter Hofstede, A. H. (2015). A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems*, *69*, 1-19.

de Leoni, M., Suriadi, S., Ter Hofstede, A. H., & Van der Aalst, W. M. (2016). Turning event logs into process movies: Animating what has really happened. *Software & Systems Modeling*, *15*(3), 707-732.

de Weerdt, J., Schupp, A., Vanderloock, A., & Baesens, B. (2013). Process mining for the multi-faceted analysis of business processes—A case study in a financial services organization. *Computers in Industry*, *64*(1), 57-67.

dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., et al. (2019). Process mining techniques and applications– A systematic mapping study. *Expert Systems with Applications*, *133*, 260-295.

Douzali, E., & Darabi, H. (2016). A case study for the application of data and process mining in intervention program assessment and improvement. In *2016 ASEE annual conference & exposition*, New Orleans, Louisiana. 10.18260/p.26267.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of Business Process Management*, 341-369, Springer Berlin, Heidelberg.

van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P. (2015). PM22: A Process Mining Project Methodology. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds) Advanced Information Systems Engineering. CAiSE 2015. Lecture Notes in Computer Science, vol 9097. Springer, Cham.

Godin, B. (2003). The impact of research grants on the productivity and quality of scientific research, *Ottawa: INRS Working Paper 2003.*

Günther, C.W., van der Aalst, W.M.P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. In: Alonso, G., Dadam, P., Rosemann, M. (eds) Business Process Management. BPM 2007. Lecture Notes in Computer Science, vol 4714. Springer, Berlin, Heidelberg.

Head, M. G., Fitchett, J. R., Atun, R., & May, R. C. (2014). Systematic analysis of funding awarded for mycology research to institutions in the UK, 1997–2010. *BMJ Open*, *4*(1), e004129.

Heyard, R., & Hottenrott, H. (2021). The value of research funding for knowledge creation and dissemination: A study of SNSF Research Grants. *Humanities and Social Sciences Communications*, *8*(1), 1-16.

Jung, H., Seo, I., Kim, J., & Kim, B. K. (2017). Factors affecting government-funded research quality. *Asian Journal of Technology Innovation*, *25*(3), 447-469.

Lakshmanan, G. T., Shamsi, D., Doganata, Y. N., Unuvar, M., & Khalaf, R. (2015). A Markov prediction model for data-driven semi-structured business processes. *Knowledge and Information Systems*, *42*(1), 97-126.

Mahmood, T., & Shaikh, G. M. (2013). Adaptive automated teller machines. *Expert Systems with Applications*, *40*(4), 1152-1169.

Mongeon, P., Brodeur, C., Beaudry, C., & Larivière, V. (2016). Concentration of research funding leads to decreasing marginal returns. *Research Evaluation*, *25*(4), 396-404.

Pagel, P. S., & Hudetz, J. A. (2015). Scholarly productivity and national institutes of health funding of foundation for anesthesia education and research grant recipients: Insights from a bibliometric analysis. *Anesthesiology, 123*(3), 683-691.

Peters, E. M., Dedene, G., & Poelmans, J. (2013). Understanding service quality and customer churn by process discovery for a multi-national banking contact center. In *IEEE 13th International Conference on Data Mining Workshops (ICDMW)*, 228-233, Texas, USA.

Ruhani A, Keshavarzi S, Anbarlou M, (2020). Providing a grounded theory of young elites' experiences in confronting bureaucracy of the National Elite Foundation. *Quarterly of Social Studies and Research in Iran, 8*(4), 745-772. In Persian.

Sonnenberg, C., & Bannert, M. (2019). Using process mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259-272.

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems*, *64*, 132-150.

Van den Beemt, A., Buijs, J., & Van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *International Review of Research in Open and Distributed Learning*, *19*(5), 38-60.

Wang, J., & Shapira, P. (2015). Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *PloS one*, *10*(2), e0117727.

Werner, M. (2017). Financial process mining-Accounting data structure dependent control flow inference. *International Journal of Accounting Information Systems*, *25*, 57-80.

Yari Eili, M., Rezaeenour, J. (2022). A survey on recommendation in process mining. *Concurrency and Computation: Practice and Experience*, *34*(26), e7304.

Zeng, Q., Sun, S. X., Duan, H., Liu, C., & Wang, H. (2013). Cross-organizational collaborative workflow mining from a multi-source log. *Decision Support Systems*, *54*(3), 1280-1301.

Appendix 1. Abbreviations, terms, and explanation of the activities

Activity ID	Abbreviation	Explanation
A	Submit	Submit request: Applicant registers through the Sina portal and uploads the requirement information
B	AutoScore	Automatic scoring: Scoring mechanism calculates the score for each applicant based on educational/ research backgrounds.
C	CentralBranch	Review by the central branch: Staff in central branch take the application to preliminary assessment. If the document needs modification, it is sent back to the applicant.
M	Reject	Rejection notification: If the score is lower than the threshold, a notification about rejection will be send to the applicant.
D	ProvincialBranch	Checked by provincial branch: After *Review by the Central Branch*, staff in provincial branch should review the application.
E	WorkingGroup	Working group reviewing: Application would be sent for review by the working group committee.
F	Modification	Information modification by applicant: Applicant would modify the documents' information.
N	Acceptance	Acceptance notification: A notification about acceptance would be sent to the applicant.
O	Canceling	Cancelling the request: The applicant withdraws the request.
G	Fund	Introducing to the financial department: The accepted application will be sent to the financial department.
H	SelectPUP	Selecting proposal topic, university, and post-doc supervisor: Applicant provides a proposal and selects the intended university and the supervisor.
I	Review / verifyPUP	Reviewing proposal, university, and post-doc supervisor: The committee would review the proposal, selected university, and the supervisor.
J	First/ second/ last quarter payment	First/ second/ last quarter payment: Funding payment through a quarter period within a year.
K	First/ second/ last quarter report	Uploading first/ second/ last quarter report: Applicants upload the scientific reports quarterly.
L	ReviewReport	First/ second/ last quarter review report: Committee reviews the scientific reports.