Supporting Decision-Making in Financial Investigations Through Prioritization of Suspicious Transactions Using Process Mining and Economic Evaluation

دعم اتخاذ القرار في التحقيقات المالية من خلال ترتيب المعاملات المشبوهة حسب الأولوية باستخدام تحليل العمليات والتقييم الاقتصادي

Mohamed S. Abu-assi

Business Information Systems Department, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt

Ibrahim E. Ragab

Department of Basic Sciences, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt

Amr M. El-seraty

Department of Economics, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt

مجلة منارة الإسكندرية للعلوم التجارية – دورية علمية محكمة

(ISSN/O : 3062 – 5548) - (ISSN/P :3062 -5696)

Supporting Decision-Making in Financial Investigations Through Prioritization of Suspicious Transactions Using Process Mining and Economic Evaluation

دعم اتخاذ القرار في التحقيقات المالية من خلال ترتيب المعاملات المشبوهة حسب الأولوية باستخدام تحليل العمليات والتقييم الاقتصادي

Mohamed S. Abu-assi[1], Ibrahim E. Ragab[2], Amr M. El-seraty[3]

[1]Business Information Systems Department, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt; mohamed.shaban@eia.edu.eg
[2]Department of Basic Sciences, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt; ibrahim.ragab@eia.edu.eg
[3]Department of Economics, Egyptian Institute of Alexandria Academy for Management and Accounting, Alexandria, Egypt; amr.elseraty@eia.edu.eg

Abstract:

This research proposes a forensic approach based on process mining to detect suspicious transactions in financial event logs. Using real-world data from the BPI Challenge 2017, the study focuses on analyzing deviations in loan offer processes recorded by a Dutch financial institution. The approach begins with preprocessing and automated discovery of process models using algorithms such as Inductive Miner, followed by conformance checking via the Multi-Perspective Process Explorer to identify deviations from expected behavior. A degree of rarity of deviation degree is then introduced to prioritize anomalies, allowing investigators to focus on the least frequent and potentially fraudulent cases. A temporal analysis using dotted charts highlights delays and irregularities in specific activities. The findings reveal deviations not only in the timing but also in the roles of originators. To support financial investigations, the study includes economic interpretation of suspicious patterns, linking anomalies to potential financial impact and systemic risk. These findings suggest that integrating degree of rarity based process mining into financial investigations significantly improves anomaly detection and resource prioritization, making it a valuable addition to institutional forensic protocols.

Keywords: Process Mining, Financial Institutions, Economic Impact, Business Process Deviation, Forensic Analysis, Fraud Detection.

المستخلص:

يقترح هذا البحث منهجًا جنائيًا يعتمد على تقنيات التنقيب في العمليات لاكتشاف المعاملات المشبوهة في سجلات الأحداث المالية. باستخدام بيانات حقيقية من تحدي BPI لعام 2017، يركز البحث على تحليل الانحرافات في عمليات عروض القروض المسجلة من قبل مؤسسة مالية هولندية. تبدأ المنهجية بمعالجة أولية واكتشاف نماذج العمليات باستخدام خوارزميات مثل Inductive Miner، يليها فحص التوافق عبر أداة Multi-Perspective Process Explorer لتحديد الانحرافات عن السلوك المتوقع. ثم يتم تقديم مقياس "درجة ندرة الانحراف" لترتيب الحالات الشاذة حسب أهميتها، مما يتيح للمحققين التركيز على أكثر الحالات ندرة وربما احتيالية. يكشف التحليل الزمني باستخدام الرسوم النقطية عن تأخيرات وعدم انتظام في أنشطة معينة. توضح النتائج وجود انحرافات في توقيت الأنشطة وفي الجهات المنفذة لها. ولتعزيز التحقيقات المالية، يتضمن البحث تحليلاً اقتصاديًا للأنماط المشبوهة، يربط بين هذه الانحرافات والتأثير المالي المحتمل والمخاطر النظامية، مما يعزز من فعالية عمليات التدقيق الجنائي في المؤسسات المالية.

الكلمات المفتاحية: التنقيب في العمليات، المؤسسات المالية، الأثر الاقتصادي، التحقيقات الجنائية، انحرافات العمليات التجارية، التحليل الجنائي، كشف الاحتيال.

1. Introduction:

Banking is a tool that allows financial institutions to bridge the gap between funding sources and execution. Banking services involve the acceptance of money deposits from the public for lending or investment purposes, which are repayable on demand or otherwise, and withdrawable by cheque, draft, or other means. Banks have introduced the modern concept of online banking, allowing customers to conduct banking activities via the Internet. These services may include various operations such as loan applications and money transfers without direct interaction with a bank employee. In fact, online banking can offer almost all traditionally provided services at local branches (Tsai et al., 2024).

In the Dutch context, electronic banking services are generally considered secure and efficient, supported by an advanced technological infrastructure and robust fraud prevention measures. According to the Dutch Central Bank (DNB), the rate of financial fraud in digital transactions remains relatively low compared to many other countries, thanks to security protocols such as two-factor authentication and advanced encryption technologies (Muhammed Raiyan A, 2024). Despite these strengths, online banking systems globally remain vulnerable to fraud due to the misuse of technology (Grammatikos & Papanikolaou, 2021; Herrera Luque et al., 2021). Misuse may include overpayments, the mishandling of sensitive information, and the improper use of digital tools. These vulnerabilities make such systems soft targets for cybercrimes and digital attacks, especially as the number of online transactions increases rapidly (Friedman, 2023).

This issue has become a global challenge, threatening the integrity and stability of financial markets and institutions (Afjal et al., 2023). According to recent Congressional reports, financial services institutions face up to 40% higher costs from cybercrime compared to other sectors (Scott & Tierno, 2023), primarily due to the volume and sensitivity of the data and transactions they handle (Grigoryan & Mirzoyan, 2023). For example, the Netherlands, while maintaining a competitive and business-friendly environment, faces challenges such as high household debt levels—reaching 103% of GDP in 2022—which pose potential risks to financial stability (Sbitenkova, 2024).

Ensuring the security and stability of banking processes is especially crucial in loan mechanisms, which enable customers to access credit and repay it with interest. Despite fraud prevention policies, internal fraud still occurs when employees exploit their access to bypass digital onboarding or manipulate internal systems to legitimize fraudulent data.

Given the limitations of human capabilities and governmental resources to thoroughly investigate cybercrimes, digital forensics (DF) has emerged as a vital solution. DF is defined as “the use of systematically derived methods to protect, gather, validate, detect, analyze, explain, and present digital evidence from digital sources with the aim of recognizing criminal activity or anticipating unauthorized actions.” Investigators must collect digital evidence to support or refute their understanding of an incident. This evidence is especially important when detecting fraud after the fact. In this regard, database forensics has become a critical subfield of DF, helping to uncover alterations in databases, even if made by authorized users (Cankaya & Kupka, 2016; Chopade & Pachghare, 2019; Olivier, 2009).

In parallel, Business Intelligence (BI) solutions have evolved to enhance end-to-end visibility of business processes. BI encompasses a variety of methods for gathering, analyzing, and monitoring business data, enabling organizations to understand how their processes function in different scenarios and to implement continuous improvements (Basile et al., 2023; Papadopoulos et al., 2018). Process Mining (PM), a powerful extension of BI, bridges the gap between data mining and process analysis. It is grounded in the extraction of event logs to construct and visualize process models for real-time monitoring and optimization (van der Aalst et al., 2012; van der Aalst et al., 2007).

This research proposes the integration of PM techniques within online banking environments to improve the detection of suspicious transactions and fraudulent activities. The framework employs process modeling to generate Petri net representations and uses conformance checking to identify the most suitable discovery algorithm for detecting suspicious behaviors.

The remainder of this paper is organized as follows: Section 2 reviews related literature. Section 3 outlines the proposed approach and methodology. Section 4 presents implementation details and experimental results, followed by a discussion. Finally, Section 5 concludes the work.

2. Literature Review:

In recent years, process mining has emerged as a powerful analytical tool for uncovering insights from event data across various domains, including healthcare (Dallagassa et al., 2022; Erdogan & Tarhan, 2018; Rojas et al., 2016; Sundari & Nayak, 2020), education (AlQaheri & Panda, 2022; Cerezo et al., 2020; Ghazal et al., 2017), cyber security (Macak et al., 2022; Mohamed & Kassem, 2023), Manufacturing (Castiglione, 2024; Céu et al., 2024), and several other sectors (Garcia et al., 2019; Maita et al., 2018). This section provides a structured review of literature related to business intelligence and economic risk detection through process mining. The reviewed works are categorized into: (1) studies on modeling business and economic workflows using process mining; (2) research integrating process mining into forensic auditing and financial investigations; and (3) studies applying process mining specifically within the banking sector to detect financial anomalies and mitigate economic risk.

2.1. Economic Workflow Modeling through Process Mining

A foundational stream of research explores how process mining can be used to model complex business operations with economic implications. (Denisov et al., 2018) explore synchronization over time by plotting case activities, offering insights into process efficiency and economic throughput. (van der Aalst, 2022) and (Burattin, 2015) take an organizational lens to identify actor-level patterns and inefficiencies that may translate into operational costs or resource misallocations.

In the healthcare domain, (Leemans et al., 2013) applied process mining to analyze patient treatment pathways, identifying bottlenecks and visualizing conformance with medical protocols—techniques that have clear parallels in economic process optimization. This work also demonstrates how α--algorithm-based discovery and visualization techniques can be translated to financial workflows.

(Rebuge & Ferreira, 2012) contribute by using sequence clustering to categorize frequent and infrequent behaviors, a method that can assist in isolating rare, high-impact economic anomalies. (van Zelst et al., 2021) compute conformance indicators from event behavior, enabling quantification of deviations that could imply economic irregularities. (Burattin & Carmona, 2018) leverage behavioral profiles to query specific event characteristics, while (Bayomie et al., 2016) apply anti-pattern recognition to flag suspicious ordering of economically relevant activities—an approach directly applicable in financial compliance auditing.

2.2. Forensics and Investigative Applications of Process Mining

Despite being underutilized in traditional database forensics, process mining is gaining traction in digital investigations involving financial evidence trails. (Englbrecht et al., 2020) reconstruct business activities from residual data on digital storage devices, showcasing the value of process mining in extracting economic actions from technical traces. (ter Voert, 2024) developed PM², a methodology tailored to apply process mining in web-based forensic investigations—providing structured guidance for investigators analyzing economic activities under legally constrained environments.

(Khan et al., 2023) introduce an approach combining rare itemset mining with process analysis to identify high-risk irregular events, ranking them based on rarity. This method is particularly effective for forensic auditors overwhelmed by high-volume transaction logs, supporting more efficient resource allocation and prioritization in economic investigations.

2.3. Process Mining in Banking: Financial Control, Risk Monitoring, and Fraud Detection

Given the volume and sensitivity of financial transactions, In fact, the banking sector is particularly suited for process mining-based audit tools. (Silva et al., 2023) model user behavior during digital onboarding using real fintech event logs, combining process mining with machine learning to achieve 80% accuracy in detecting fraudulent users. Their findings underscore the potential of hybrid systems in identifying financial fraud risks during customer acquisition.

(Jans et al., 2014) conducted one of the earliest studies to apply process mining for banking audits using data from a leading European bank. Their work exposed unauthorized payments and internal control violations, demonstrating how automated log analysis can reveal economic threats previously undetected by manual audits. (Chiu & Jans, 2019) build on this by identifying weak points in internal financial controls using event log auditing.

(Broer Bahaweres et al., 2021) show how ProM tools can be effectively deployed in finance-related logs to detect anomalies, reduce inefficiencies, and ultimately mitigate financial risk exposure. Meanwhile, (Rodríguez-Quintero et al., 2021) highlight how visual analytics, such as dotted charts and trace alignments, enhance fraud detection in audit processes. Despite its effectiveness, their approach faced limitations due to manual event handling—gaps that the current research addresses by introducing automated deviation ranking and report generation for financial investigators.

While previous studies provide significant insights into process mining applications, several limitations remain unaddressed—particularly in post-fraud analysis and structured reporting for financial investigators. Existing works often lack a standardized approach for ranking deviations based on economic rarity or for translating process anomalies into actionable forensic reports. Moreover, while visual tools and conformance checking techniques have been successfully employed in other sectors, few have been adapted to the specific needs of financial institutions, where traceability, documentation, and legal accountability are critical.

The current study addresses these gaps by proposing a business intelligence framework based on process mining tailored to post-fraud financial investigations. It integrates multiple discovery algorithms and conformance checking tools to identify unusual process variants, computes a rarity-based deviation score for prioritization, and generates structured reports including economically meaningful activities (e.g., “O_Cancelled,” “O_Refused,” “O_Accepted”) for forensic analysts. This contributes not only to the process mining field but also to the broader discipline of economic and financial crime investigation by enhancing detection accuracy, reducing investigative overload, and increasing audit transparency.

3. Methodology:

The techniques of process mining are used mainly to analyze business processes based on event logs that are often extracted from database systems

(González López de Murillas et al., 2019). Recently, process mining techniques are increasingly applied in various domains, including digital forensics to uncover insights about process execution and compliance (Englbrecht et al., 2020; Khan et al., 2023; ter Voert, 2024).

In fact, the main contribution of the current work is to introduce an intelligent analytic framework for forensic applications to assist in discovering all suspicious behaviors in bank sector. The proposed approach is based on the "Fraud data analytics methodology" by (Vona, 2016) to discover all suspicious behavior in the bank systems by analyzing data for red flags that related to specific suspicious scenario. This methodology is considered as an internationally fundamental in the field of fraud analysis (Rodríguez-Quintero et al., 2021).

The goal is to implement the process mining techniques to support the specialized forensic applications for detect suspicious transactions and fraudulent activities for a purpose of assisting forensic investigators by reducing the required workload and effort required compared to traditional analysis.

In general, the process mining techniques help to model and analyze the business process. The novelty of the current work is the employing of process mining for discovering specific suspicious behaviors from event log files. In this sense, the proposed approach aims at better tracing deviations in complex datasets and providing investigators with a refined tool for reconstructing transaction sequences post-incident.

Unlike related works that rely on manual data handling or narrow data scopes, proposed approach aims to integrate multi-perspective analysis with robust process mining techniques to improve the efficiency and accuracy of suspicious detection across various systems, automates the prioritization of suspicious activities, and generates comprehensive reports to aid investigators in decision-making.

3.1 The proposed approach

In process mining, the event log file serves as the fundamental input for all techniques. It comprises multiple cases, each representing a sequence of events uniquely linked to that case. These events are chronologically ordered based on their timestamps. Each event may contain attributes such as the activity name, timestamp, associated costs, and the resource involved. The ordered sequence of events within a case is referred to as a trace. The proposed approach is designed to improve both the effectiveness and efficiency of forensic investigations, as shown in Figure 1.

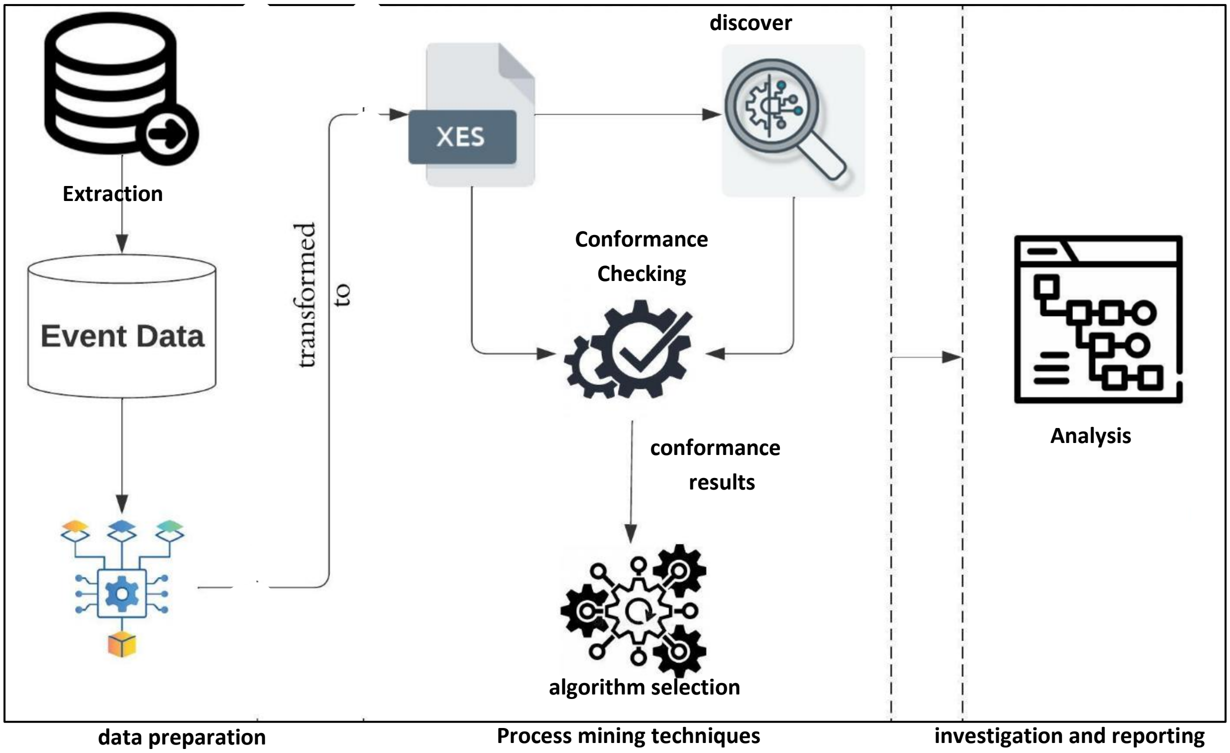


Fig. 1. Conceptual diagram of proposed approach.

In the context of analyzing business processes, event data refers to a structured collection of records generated during the execution of various activities. Each record, or event, is typically represented by a tuple such as $(CaseID\ ,T,TS,..)$. Here, CaseID uniquely identifies a specific instance of the process, T denotes the type of activity or event performed, and TS captures the timestamp indicating when the event occurred. Additional attributes may also be associated with each event, such as resource identifiers, cost elements, or organizational roles, offering a richer context for process analysis.

An event corresponds to a single execution step within a business process. Each event carries a set of attributes—including activity name, timestamp, and involved resource—that provide descriptive and analytical value. Let Σ represent the universe of all possible events. For any event $e \in \Sigma$, the value of a particular attribute A can be accessed via a function denoted $\#_A\ (\ e\ )$. A process instance, also referred to as a case, is defined by a unique identifier and consists of a sequence of temporally ordered events linked to the execution of a specific scenario within the broader process. Collectively, these instances are captured in what is known as an event log, typically represented as a set $L\ \subseteq\ \Sigma^{\ *}$. Each element of L, called a trace, reflects the full chronological sequence of events for one case and illustrates a particular path or variant through the process (Silva et al., 2023).

A trace, in this context, refers to a non-empty and strictly ordered sequence of events associated with a single case. It documents the flow of activity executions from start to end and serves as a basis for identifying different behavioral patterns or deviations across multiple instances. In many process mining applications, the name of the activity itself is often used as a simplified representation of the event, facilitating the construction and analysis of sequences such as (a,b,d), which indicates a three-step activity flow (Silva et al., 2023).

To model the control-flow perspective of such processes, Petri nets are widely used due to their formal semantics and graphical representation capabilities. A Petri net is composed of places (representing conditions or states), transitions (representing activities or events), and flow relations that define how tokens move between places and transitions. These relations are expressed through directed arcs and ensure the logical structure of the process model. The state of the Petri net at any given time is defined by a marking, which specifies the distribution of tokens across places. This marking governs which transitions are enabled and hence which parts of the process can proceed. A marked Petri net combines the static structure of the net with its current state and serves as a foundational tool for conformance checking and behavioral analysis in process mining studies.

The quality of the event log plays a critical role in shaping the accuracy and interpretability of the resulting business process model. However, event logs often suffer from inherent quality issues due to the presence of deviations, noise, and inconsistencies. These irregularities can lead to the generation of highly complex and entangled models—commonly referred to as "spaghetti models"—which are difficult to interpret and fail to represent the actual behavior of business processes. The excessive interconnection of activities within such models obscures the underlying process logic and reduces their analytical value.

To mitigate this challenge, event log filtering and preprocessing are essential preparatory steps. These techniques help reduce complexity and enhance the clarity of discovered models, enabling a more accurate depiction of genuine business behavior. Once the event logs are cleaned and structured, process mining techniques can be applied effectively to uncover typical activity sequences and identify deviations from expected workflows.

In the proposed approach, three foundational process mining techniques are integrated to support the discovery of business process patterns and potentially suspicious behaviors. **First**, automated process discovery methods are used to generate process models from the event logs, capturing all possible execution variants. This transformation provides a structural view of how activities are sequenced and executed in practice, as illustrated in Figure 2.

Second, conformance checking is employed to assess the alignment between the observed event data and the discovered models. By comparing the recorded behavior with the expected process flows, this technique helps identify discrepancies that may indicate irregularities or non-compliance. In this study, conformance checking is performed using the "Multi-Perspective Process Explorer" in the ProM framework, where event logs are analyzed against multiple Petri nets to highlight deviations and detect potentially suspicious actions.

Finally, the selection of the optimal algorithm for conformance and deviation analysis is based on evaluation metrics such as precision, which assesses how well the model excludes irrelevant behavior, and fitness (or recall), which measures how accurately the model reflects the observed data. These metrics support the extraction and prioritization of deviating variants, enabling investigators to focus on cases that warrant further forensic or financial analysis.

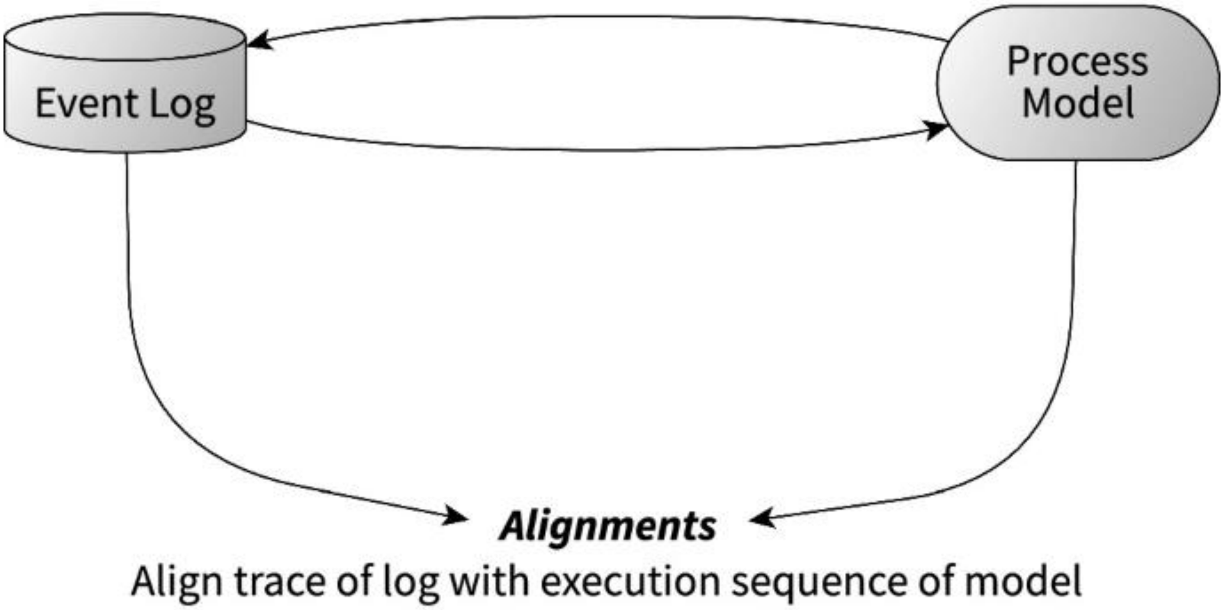


Fig. 2. Applying process mining techniques, specifically conformance checking, as a methodology for detecting suspicious activities.

In the final phase of the proposed approach, each identified deviation undergoes further analysis to introduce a quantitative degree that prioritizes anomalies for investigation. This prioritization mechanism enables investigators to allocate their attention and resources more effectively by focusing on the most significant and potentially suspicious deviations first. To support this, a comprehensive reference document is generated for each case flagged as suspicious. These documents summarize the nature of the deviations, highlight the relevant findings, and serve as a valuable resource to facilitate informed decision-making during the investigation process.

The analysis begins by examining the results obtained from the conformance checking phase. Using the "Multi-Perspective Process Explorer" tool in the ProM framework, the approach identifies variants that significantly deviate from the expected process model or match predefined patterns of suspicious behavior. These deviant variants are scrutinized further to detect and isolate transactions that may be indicative of fraudulent activity.

The key component in this analysis is the introduction of the Degree of Rarity of Deviation, a degree designed to assess the significance of each suspicious variant based on its frequency in the event log. This measure is calculated by comparing the number of cases associated with a specific variant to the total number of cases in the dataset. The underlying assumption is that rare variants are more likely to correspond to unusual or unauthorized process executions. By prioritizing the least common deviations, this step enhances the model's capacity to detect hidden or subtle instances of fraud. The calculation and evaluation of Degree of Rarity are formally outlined in Algorithm 1.

Algorithm 1: Rarity Score Computation for Suspicious Variants

```
Input: suspiciousVariants, totalEventsXESFile
Output: degreeOfRarity

1. Determine the total number of cases (totalCasesCount) from the event
   log (totalEventsXESFile).
2. Create an empty collection named rarityScores.
3. Iterate through each suspicious variant in suspiciousVariants:
      a. Obtain the number of cases associated with the current variant
         (variantCaseCount).
      b. Compute the rarity percentage using the formula:
            $rarity = (variantCaseCount / totalCasesCount) \times 100$.
      c. Add a record consisting of the variant, its case count, and the
         computed rarity to rarityScores.
4. Organize the rarityScores list in ascending order based on the rarity
   value.
5. Present the results and export them as needed.

Return rarityScores.
```

4. Implementation and Experimental Results:

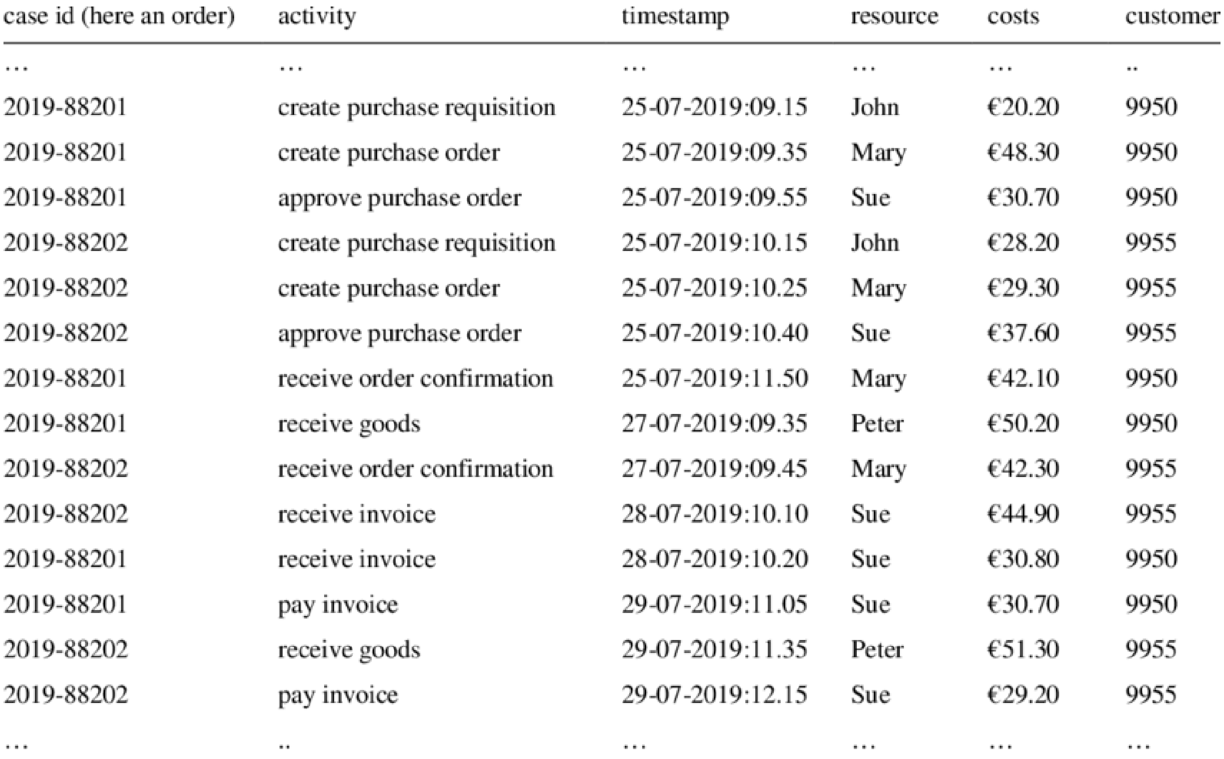
The implementation of the proposed approach is based on a series of experiments that are conducted using a laptop in a Windows environment using ProM version 6.13 (Ramos & Rossi, 2023).

4.1 Event Log Extraction

The primary input for the proposed approach is a real-world event log sourced from the Business Process Intelligence (BPI) Challenge 2017, specifically the Offer Event Log. This log captures traces of loan offer operations conducted by a Dutch financial institution. The dataset was originally extracted from an internal database and subsequently converted into a structured event log format suitable for process mining analysis. It documents the execution of a loan application process carried out in 2016, encompassing 1,202,267 recorded events linked to 31,509 individual loan applications.

Among these applications, a total of 42,995 loan offers were generated. The subset of the event log relevant to this study focuses on the offer management process, comprising 193,849 events distributed across the 42,995 offer cases. These events correspond to eight distinct activity types: O_Create offer, O_Created, O_Sent (online only), O_Sent (mail and online), O_Returned, O_Accepted, O_Cancelled, and O_Refused. Each offer is intrinsically associated with a corresponding loan request, marking the beginning of the process when a customer submits a loan application. As the process unfolds, the offer is iteratively updated through ongoing interactions between the client and the financial institution, culminating in either acceptance or rejection. In this context, an offer represents a complete instance of the business process being analyzed.

Event data in this log is structured in a straightforward attribute-value format, which allows for easy representation in a two-dimensional table. This format facilitates the application of process mining techniques. Figure 3 illustrates the general structure of the event logs utilized in this study to extract process behavior and analyze deviations effectively.



case id (here an order)	activity	timestamp	resource	costs	customer
…	…	…	…	…	..
2019-88201	create purchase requisition	25-07-2019:09.15	John	€20.20	9950
2019-88201	create purchase order	25-07-2019:09.35	Mary	€48.30	9950
2019-88201	approve purchase order	25-07-2019:09.55	Sue	€30.70	9950
2019-88202	create purchase requisition	25-07-2019:10.15	John	€28.20	9955
2019-88202	create purchase order	25-07-2019:10.25	Mary	€29.30	9955
2019-88202	approve purchase order	25-07-2019:10.40	Sue	€37.60	9955
2019-88201	receive order confirmation	25-07-2019:11.50	Mary	€42.10	9950
2019-88201	receive goods	27-07-2019:09.35	Peter	€50.20	9950
2019-88202	receive order confirmation	27-07-2019:09.45	Mary	€42.30	9955
2019-88202	receive invoice	28-07-2019:10.10	Sue	€44.90	9955
2019-88201	receive invoice	28-07-2019:10.20	Sue	€30.80	9950
2019-88201	pay invoice	29-07-2019:11.05	Sue	€30.70	9950
2019-88202	receive goods	29-07-2019:11.35	Peter	€51.30	9955
2019-88202	pay invoice	29-07-2019:12.15	Sue	€29.20	9955
…	..	…	…	…	…

Fig. 3. A structure for an event log

Event data is usually structured in an attribute-value format, easily represented in a 2-dimensional table. To unify the input format of process mining, the XML-based MXML (Mining XML) format and its meta-model were introduced (van Dongen & Van der Aalst, 2005). However, due to limitations in MXML, the XES (eXtensible Event Stream) standard emerged under IEEE to enhance flexibility and interoperability (Verbeek et al., 2011). In practice, proposed approach leverages the XES format by converting raw event logs using tools like the Convert CSV to XES plug-in in ProM. The framework assumes that the bank's information system is designed to store process traces in the XES format, enabling seamless analysis. Once process instances (e.g., loan offers) are completed, their events are systematically transferred to a centralized event log file (Acampora et al., 2017). the processed data is exported into the XES format, creating a standardized event log ready for advanced process mining techniques.

4.2 Process mining techniques

Business process modeling always starts from event logs using process discovery techniques to obtain process models in terms of a Petri net. In this work, four popular miners are applied including Alpha Miner, Inductive Miner, ILP Miner, and Heuristics Miner to generate the business process models.

Alpha Miner: An alpha miner (Aalst et al., 2004) is a popular process mining technique for returning interconnection of events. It is used to generate the process model in a form of workflow nets which are special class of the Petri net. However, this miner suffers from many problems especially when deal with noisy data, duplicated and hidden tasks.

Heuristics Miner: A (Weijters & Ribeiro, 2011) can be seen as an enhanced version of alpha miner as it considers frequency and has an ability to work with short loops. The visual model by heuristics miner represents events as nodes that are interconnected by directed edges with frequency and performance calculations. The frequency signifies the number of times the source event follows the destination event. However, the heuristics miner is limited in declaration the soundness of the process model.

Inductive Miner: An Inductive miner (Leemans et al., 2013) is characteristic by its ability to deal with huge event logs and its ability to ensure a soundness of the process model. In addition, it has an ability to deal with hidden tasks, noisy and incomplete data.

The Multi-perspective Process Explorer framework is used to evaluate the algorithms' suitability for process discovery. Four key metrics were analyzed: Average Activity Precision, Average Fitness, Percentage of Violations, and Events (Correct, Wrong, Missing), providing insights into the algorithms' capabilities and accuracy.

4.3 Comparative Analysis Results

Presents a comparative analysis of process discovery algorithms based on their ability to reconstruct the offer lifecycle from event logs. The evaluation considers precision, fitness, F1-score and execution time (in seconds, measured randomly) as key metrics for assessing the discovered models. Figure 4 present the detailed results.

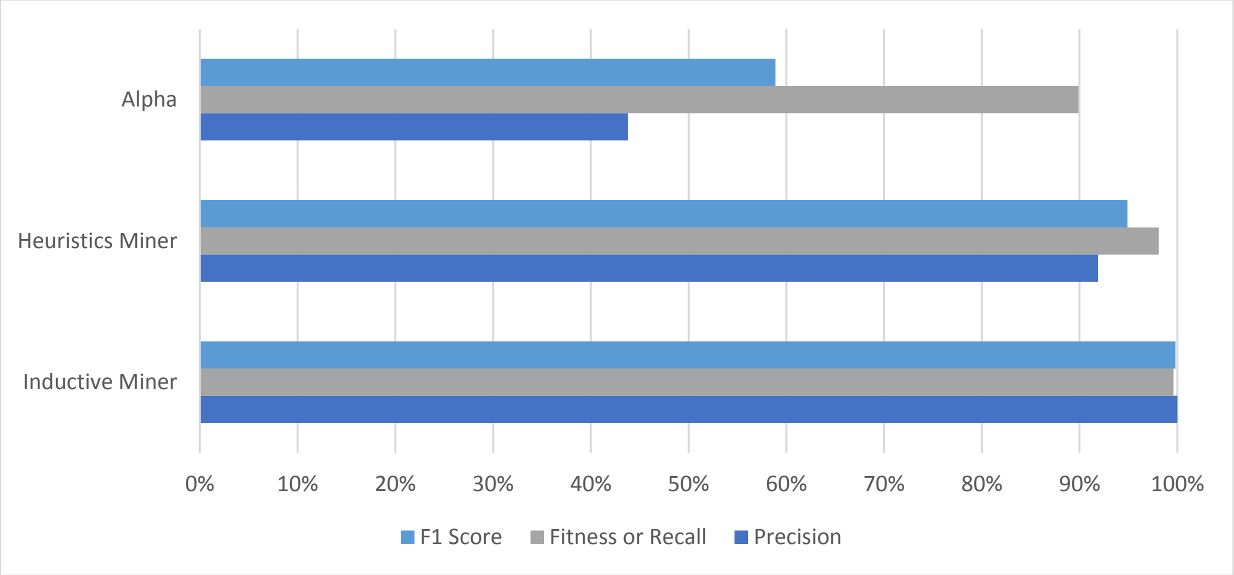


Fig. 4. A comparative analysis of the overall performance metrics of the algorithms.

The evaluation of multiple process discovery algorithms revealed that **the Inductive Miner** consistently outperformed its counterparts across all performance metrics. It achieved perfect precision (100%), a fitness score of 99.6%, and an F1 score of 99.8%, clearly demonstrating its robustness and high accuracy in capturing real process behavior. Notably, it processed the entire set of 193,849 events correctly, with minimal deviations, indicating excellent alignment between the discovered model and the actual event log.

The Heuristics Miner also performed relatively well, achieving a precision of 91.9%, fitness of 98.1%, and an F1 score of 94.9%. While it demonstrated a balanced performance, certain complex or infrequent behaviors were not as well-fitted as those captured by the Inductive Miner, slightly affecting its overall reliability.

By contrast, **the Alpha Miner** exhibited the weakest performance, with a precision of only 43.8%, fitness of 89.9%, and an F1 score of 58.9%. These results highlight its limitations in handling complex or realistic event logs, as it tends to produce oversimplified models that fail to represent the true nature of the process.

Overall, the experimental results firmly establish the Inductive Miner as the most reliable and effective algorithm for generating accurate and comprehensive process models. Its ability to closely match observed behaviors recorded in the event log makes it particularly suitable for forensic analysis and anomaly detection in business process environments.

In the analyzed event log, eight distinct activities are associated with the lifecycle of a credit offer. The process begins with “O_Create offer”, representing the creation of a credit offer. It proceeds through subsequent stages such as “O_Sent (online only)”, “O_Sent (mail and online)”, and “O_Returned”, reflecting the communication and document return steps between the client and the financial institution. The later stages confirm the final status of the offer, including “O_Accepted” (indicating approval), “O_Cancelled” (offer cancelled by the client), and “O_Refused” (offer rejected by the bank). Together, these activities cover the complete lifecycle of credit offer processing within the system, forming the basis for understanding typical and atypical process flows.

4.4 Timing Analysis Results

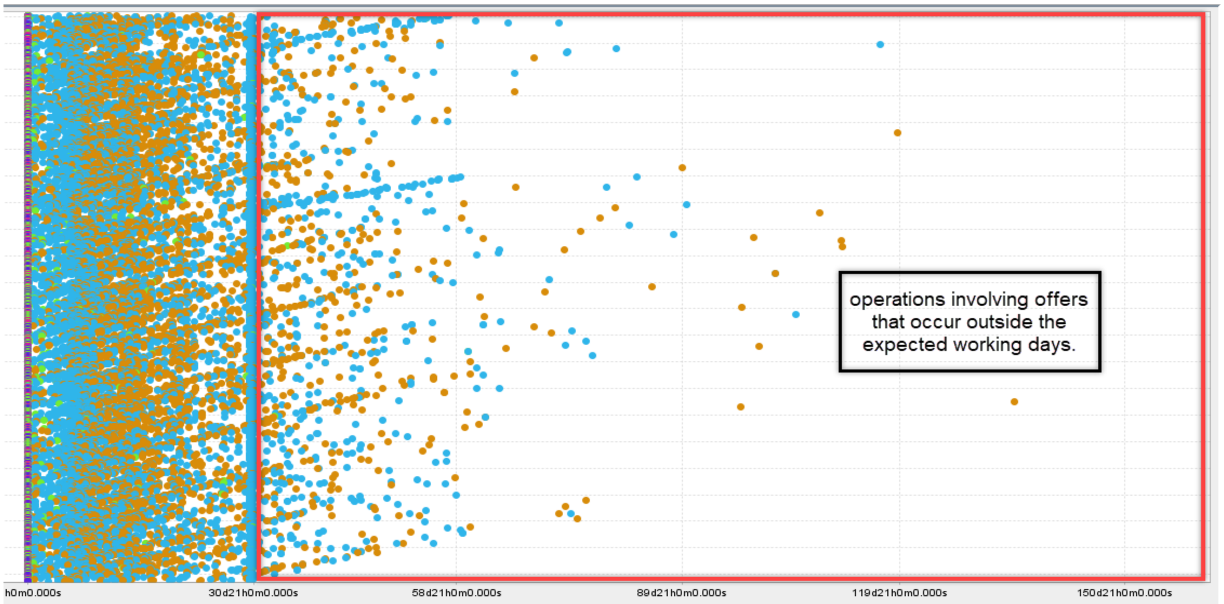


Figure 5: View of instance duration represented through a dotted chart.

To further investigate the temporal behavior of these process instances, a dotted chart was employed. This visualization technique maps each event as an individual point, where the color and shape of the points encode specific event attributes, such as activity type or resource. Figure 5 presents this temporal distribution, illustrating the sequence of operations performed on each offer (displayed along the rows) across time (represented on the x-axis in days). This visual approach facilitates the identification of irregular patterns, bottlenecks, or anomalies in process execution.

The chart illustrates a dense cluster of points (events) corresponding to offers processed within the first month of their creation. This aligns with the expected duration that an offer typically remains active in the system. However, the red box highlights a notable number of outlier points, suggesting that multiple operations were performed on offers exceeding a month. This behavior warrants attention, as the analysis revealed that the average time for offers to remain in the system is approximately one month. Such deviations may indicate unusual or extended processing, which could be attributed to exceptions or anomalies in the loan application process.

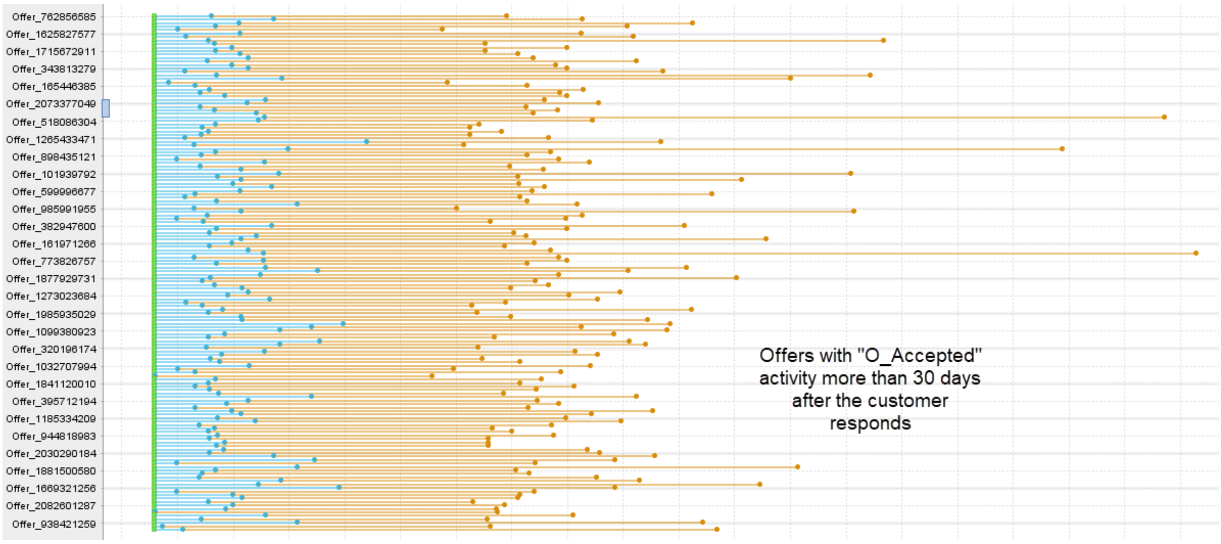


Figure 6: Viewing the operations for offers with "O_Accepted" activity more than 30 days after customer response.

Building on these insights, Figure 6 provides a complementary logical view using a dotted chart. In this chart, each row corresponds to an individual offer, with colored dots representing the performed activities and the x-axis indicating the time intervals between them. The analysis revealed 148 cases in which acceptance actions (orange dots) were recorded more than 30 days after the customer response (light blue dots). These extended durations suggest potential deviations from standard procedures. Notably, several originators were consistently involved in these delayed actions.

As a result of this temporal analysis, the phase concluded with the identification of users who frequently executed activities beyond the expected operational timeframe. Moreover, the findings revealed a disconnect between the users responsible for recording the "O_Accepted" activity and those who logged the client's response to the offer. This discrepancy highlights potential process irregularities and provides a more comprehensive understanding of deviations within the workflow, particularly regarding the distinct roles of users involved in managing and finalizing credit offers.

4.5 Deviation Prioritization and Fraud Risk Assessment

Accordingly, the computed rarity scores provide a quantitative indicator of how frequently each suspicious variant occurs within the overall event log. This degree plays a critical role in prioritizing investigative efforts by highlighting the least frequent — and potentially most abnormal — behavior patterns. By integrating this rarity-based measure into the fraud detection framework, the analysis enables investigators to concentrate first on the most exceptional and atypical process variants, as summarized in Table 1. This targeted approach significantly enhances the efficiency of forensic analysis by ensuring that attention and resources are allocated to the cases most likely to reveal fraudulent or anomalous activity.

Table 1. Investigation Priority Table Based on Rarity.

#	Cases	Sequence of events	Degree of Rarity
Variant 1	146	O_Create Offer → O_Created → O_Sent (mail and online) → O_Returned → O_Accepted	0.34 %
Variant 2	2	O_Create Offer → O_Created → O_Sent (online only) → O_Returned → O_Accepted	0.004 %
Variant 3	21	O_Create Offer → O_Created → O_Sent (mail and online) → O_Returned → O_Refused	0.049 %
Variant 4	241	O_Create Offer → O_Created → O_Sent (mail and online) → O_Returned → O_Cancelled	0.56 %

Variant 5	6	O_Create Offer → O_Created → O_Sent (online only) → O_Returned → O_Cancelled	0.014 %

The five identified process variants reflect differing degrees of procedural completeness, control effectiveness, and data integrity within the credit offer lifecycle, each carrying distinct economic implications.

Variant 1, although comparatively more frequent, still accounts for a small proportion of total cases (146 occurrences, 0.34%). It involves the creation, dual-channel dispatch, return, and final acceptance of the credit offer. Despite being a seemingly successful and complete process, the rarity of this sequence suggests that such optimal workflows are underutilized. Economically, this underuse may indicate inefficiencies in channel integration or client engagement. From a macroeconomic perspective, it suggests that investments in multichannel digital infrastructure are not yielding proportional returns, thereby limiting the financial sector's potential to drive digital inclusion, expand access to credit, and stimulate broader economic modernization.

Variant 2, observed in only two cases (0.004%), follows a similar structure to the first but limits communication to online channels. Its extreme rarity may reflect system-level anomalies or isolated instances of non-standard behavior, potentially indicative of internal testing or controlled attempts to bypass traditional communication flows. From an economic standpoint, this pattern underscores concerns around digital system vulnerabilities. Such infrequent but complete digital-only interactions may signal potential weaknesses in online identity verification or process security. These weaknesses elevate institutional cybersecurity risks, necessitating higher public and private spending on digital safeguards. At scale, this dynamic contributes to increased fiscal burdens, reduced trust in fintech ecosystems, and diminished appeal for foreign direct investment (FDI).

Variant 3, consisting of 21 cases (0.049%), captures a scenario where the offer is returned but subsequently refused. While the client appears to have fulfilled their role, the decision to reject post-submission may suggest opaque internal criteria, conflicting decision rules, or process bottlenecks. Economically, this pattern introduces inefficiencies in the credit pipeline, incurring administrative costs without financial return. On a broader scale, such inconsistencies distort credit demand signals, undermine the transparency of loan approval systems, and potentially exclude creditworthy individuals from accessing capital. This misallocation of credit hampers entrepreneurship reduces financial inclusion, and ultimately slows down private sector growth.

Variant 4—the most frequent among the five (241 cases, 0.56%)—involves a similar path, ending with cancellation rather than acceptance or rejection. While client participation is evident, the process termination at this stage may result from dissatisfaction, lack of responsiveness, or changing borrower intentions. Although more common, the large volume of such incomplete transactions reflects a structural inefficiency in the client-bank interaction model. Economically, it implies a recurring waste of institutional resources and contributes to inflated application figures that do not translate into credit disbursement. At the macro level, this scenario creates reporting inconsistencies and overstates actual credit uptake, potentially misleading policy decisions and hindering effective monetary targeting.

Variant 5, recorded in just six instances (0.014%), mirrors the structure of variant 4 but is conducted exclusively via online channels. Its rarity suggests abnormal client behavior or possibly non-genuine applications, raising suspicions of system misuse through dummy accounts or automated scripts. Such entries compromise the reliability of financial data and pollute key indicators like loan origination rates and credit conversion ratios. Over time, these distortions can skew non-performing loan (NPL) metrics and increase the burden on capital adequacy buffers. Consequently, banks may adopt more conservative lending practices, tightening credit conditions and dampening private investment, which negatively affects long-term economic dynamism and credit market efficiency.

Unlike prior works using static rule-based detection or ML classifiers, our approach emphasizes traceability, transparency, and investigator-centric prioritization through degree of rarity scores. Future work may compare the performance of our method against supervised anomaly detection benchmarks.

5. Conclusion and Future Work:

This research presented a comprehensive study on the effectiveness of applying advanced process mining techniques in forensic applications of databases to combat digital crimes and examining their implications for macroeconomic indicators. The key findings demonstrated that the proposed approach, implemented using the ProM tool, is capable of efficiently and accurately detecting suspicious fraudulent behaviors in the banking sector, outperforming traditional methods. The core contribution lies in providing an integrated methodology encompassing process modeling and model analysis, which has been practically validated in a vital sector.

Despite these promising results, future work can be developed in several directions. These include expanding the scope of application to other sectors, enhancing the methodology by integrating machine learning techniques, exploring other types of digital crimes, developing specialized visual analytics tools, and conducting further empirical studies in diverse environments.

Future research will explore dynamic variants of the degree of rarity of deviation degree, suitable for integration with real-time stream processing engines such as Apromore or PM4Py streaming modules.

Data Availability Statement

The dataset used in this study was sourced from the BPI Challenge 2017 (available at: https://data.4tu.nl/articles/dataset/BPI_Challenge_2017/12696884; accessed on 31 March 2025).

References

Aalst, W. v. d., Weijters, T., & Maruster, L. (2004). Workflow mining: discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128-1142. https://doi.org/10.1109/TKDE.2004.47

Acampora, G., Vitiello, A., Di Stefano, B., van der Aalst, W., Günther, C., & Verbeek, E. (2017). Ieee 1849tm: The xes standard. *IEEE Computational Intelligence Magazine*, 4-8.

Afjal, M., Salamzadeh, A., & Dana, L.-P. (2023). Financial Fraud and Credit Risk: Illicit Practices and Their Impact on Banking Stability. *Journal of Risk and Financial Management*, *16*(9), 386. https://www.mdpi.com/1911-8074/16/9/386

AlQaheri, H., & Panda, M. (2022). An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality. *Information*, *13*(1), 29. https://www.mdpi.com/2078-2489/13/1/29

Basile, L. J., Carbonara, N., Pellegrino, R., & Panniello, U. (2023). Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making. *Technovation*, *120*, 102482. https://doi.org/https://doi.org/10.1016/j.technovation.2022.102482

Bayomie, D., Awad, A., & Ezat, E. (2016). Correlating Unlabeled Events from Cyclic Business Processes Execution. In S. Nurcan, P. Soffer, M. Bajec, & J. Eder, *Advanced Information Systems Engineering* Cham.

Broer Bahaweres, R., Trawally, J., Hermadi, I., & Imam Suroso, A. (2021). Forensic Audit Using Process Mining to Detect Fraud. *Journal of Physics: Conference Series*, *1779*(1), 012013. https://doi.org/10.1088/1742-6596/1779/1/012013

Burattin, A. (2015). Process mining techniques in business environments. *Lecture Notes in Business Information Processing*, *207*, 220.

Burattin, A., & Carmona, J. (2018). A Framework for Online Conformance Checking. In E. Teniente & M. Weidlich, *Business Process Management Workshops* Cham.

Cankaya, E. C., & Kupka, B. (2016, 6-7 Dec. 2016). A survey of digital forensics tools for database extraction. 2016 Future Technologies Conference (FTC),

Castiglione, C. (2024). Automated generation of digital models for manufacturing systems: The event-centric process mining approach. *Computers & Industrial*

Engineering, *197*, 110596. https://doi.org/https://doi.org/10.1016/j.cie.2024.110596

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74-88. https://doi.org/10.1007/s12528-019-09225-y

Céu, H., Grilo, C., Rijo, R., & Martinho, R. (2024). Mining Resource Usage in Molds Manufacturing Processes through Process Mining. *Procedia Computer Science*, *239*, 2359-2368. https://doi.org/https://doi.org/10.1016/j.procs.2024.06.429

Chiu, T., & Jans, M. (2019). Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness. *Accounting Horizons*, *33*(3), 141-156. https://doi.org/10.2308/acch-52458

Chopade, R., & Pachghare, V. K. (2019). Ten years of critical review on database forensics research. *Digital Investigation*, *29*, 180-197. https://doi.org/https://doi.org/10.1016/j.diin.2019.04.001

Dallagassa, M. R., dos Santos Garcia, C., Scalabrin, E. E., Ioshii, S. O., & Carvalho, D. R. (2022). Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. *Journal of Ambient Intelligence and Humanized Computing*, *13*(1), 165-182. https://doi.org/10.1007/s12652-021-02894-7

Denisov, V., Fahland, D., & van der Aalst, W. M. P. (2018). Unbiased, Fine-Grained Description of Processes Performance from Event Data. In M. Weske, M. Montali, I. Weber, & J. vom Brocke, *Business Process Management* Cham.

Englbrecht, L., Schönig, S., & Pernul, G. (2020). Supporting Process Mining with Recovered Residual Data. In J. Grabis & D. Bork, *The Practice of Enterprise Modeling* Cham.

Erdogan, T. G., & Tarhan, A. (2018). Systematic Mapping of Process Mining Studies in Healthcare. *IEEE Access*, *6*, 24543-24567. https://doi.org/10.1109/ACCESS.2018.2831244

Friedman, A. (2023). Digital Economy and Place-Making. In A. Friedman (Ed.), *The Sustainable Digital City* (pp. 159-168). Springer International Publishing. https://doi.org/10.1007/978-3-031-25488-8_10

Garcia, C. d. S., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., & Scalabrin, E. E. (2019). Process mining techniques and applications – A systematic mapping study. *Expert Systems with Applications*, *133*, 260-295. https://doi.org/https://doi.org/10.1016/j.eswa.2019.05.003

Ghazal, M. A., Ibrahim, O., & Salama, M. A. (2017, 17-19 Nov. 2017). Educational Process Mining: A Systematic Literature Review. 2017 European Conference on Electrical Engineering and Computer Science (EECS),

González López de Murillas, E., Reijers, H. A., & van der Aalst, W. M. P. (2019). Connecting databases with process mining: a meta model and toolset. *Software & Systems Modeling*, *18*(2), 1209-1247. https://doi.org/10.1007/s10270-018-0664-7

Grammatikos, T., & Papanikolaou, N. I. (2021). Applying Benford's Law to Detect Accounting Data Manipulation in the Banking Industry. *Journal of Financial Services Research*, *59*(1), 115-142. https://doi.org/10.1007/s10693-020-00334-9

Grigoryan, L., & Mirzoyan, L. (2023). Cybersecurity Risks and Its Regulations. The Philosophy of Cybersecurity Audit. *WISDOM*, *25*(1), 67-77. https://doi.org/10.24234/wisdom.v25i1.970

Herrera Luque, F. J., Munera López, J., & Williams, P. (2021). Cyber risk as a threat to financial stability. *Revista de Estabilidad Financiera/Banco de España, 40 (primavera 2021), p. 181-205*.

Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A Field Study on the Use of Process Mining of Event Logs as an Analytical Procedure in Auditing. *The Accounting Review*, *89*(5), 1751-1773. https://doi.org/10.2308/accr-50807

Khan, S., Parkinson, S., & Murphy, C. (2023). Context-based irregular activity detection in event logs for forensic investigations: An itemset mining approach. *Expert Systems with Applications*, *233*, 120991. https://doi.org/https://doi.org/10.1016/j.eswa.2023.120991

Leemans, S. J. J., Fahland, D., & van der Aalst, W. M. P. (2013). Discovering Block-Structured Process Models from Event Logs - A Constructive Approach. In J.-M. Colom & J. Desel, *Application and Theory of Petri Nets and Concurrency* Berlin, Heidelberg.

Macak, M., Daubner, L., Fani Sani, M., & Buhnova, B. (2022). Process mining usage in cybersecurity and software reliability analysis: A systematic literature review. *Array*, *13*, 100120. https://doi.org/https://doi.org/10.1016/j.array.2021.100120

Maita, A. R. C., Martins, L. C., López Paz, C. R., Rafferty, L., Hung, P. C. K., Peres, S. M., & Fantinato, M. (2018). A systematic mapping study of process mining. *Enterprise Information Systems*, *12*(5), 505-549. https://doi.org/10.1080/17517575.2017.1402371

Mohamed, R. A., & Kassem, G. (2023, 22-24 Oct. 2023). Development of Conceptual Model for Performing Process Mining on Blockchain Data: A Cybersecurity Approach. 2023 2nd International Conference on Smart Cities 4.0,

Muhammed Raiyan A, S. C., Harikrishnan P R, Kavya Mohan K, Bhavana S, Sanjana Varma, Dr. Gobi Natesan. (2024). PREVENTION OF CYBER FRAUDS IN THE BANKING SECTOR. *International Scientific Journal of Engineering and Management*

03(03). https://doi.org/10.55041/ISJEM01341

Olivier, M. S. (2009). On metadata context in Database Forensics. *Digital Investigation*, *5*(3), 115-123. https://doi.org/https://doi.org/10.1016/j.diin.2008.10.001

Papadopoulos, G. A., Kechagias, E., Legga, P., & Tatsiopoulos, I. (2018). Integrating business process management with public sector. Proceedings of the international conference on industrial engineering and operations management,

Ramos, E. O., & Rossi, R. (2023). Process Mining Applied in a Software Project Development with SCRUM and ProM. *European Journal of Engineering and Technology Research*, *8*(5), 17-24. https://doi.org/10.24018/ejeng.2023.8.5.3089

Rebuge, Á., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, *37*(2), 99-116. https://doi.org/https://doi.org/10.1016/j.is.2011.01.003

Rodríguez-Quintero, J.-F., Sánchez-Díaz, A., Iriarte-Navarro, L., Maté, A., Marco-Such, M., & Trujillo, J. (2021). Fraud Audit Based on Visual Analysis: A Process Mining Approach. *Applied Sciences*, *11*(11), 4751. https://www.mdpi.com/2076-3417/11/11/4751

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, *61*, 224-236. https://doi.org/https://doi.org/10.1016/j.jbi.2016.04.007

Sbitenkova, V. M. (2024). INVESTMENT ENVIRONMENT OF THE NETHERLANDS. *European Vector of Economic Development.*, *1(36)*. https://doi.org/https://doi.org/10.32342/2074-5362-2024-1-36-7

Scott, A. P., & Tierno, P. (2023). *Introduction to Financial Services: Financial Cybersecurity*. https://crsreports.congress.gov/product/pdf/IF/IF11717

Silva, M. C. d., Tavares, G. M., Gritti, M. C., Ceravolo, P., & Barbon Junior, S. (2023). Using Process Mining to Reduce Fraud in Digital Onboarding. *FinTech*, *2*(1), 120-137. https://www.mdpi.com/2674-1032/2/1/9

Sundari, M. S., & Nayak, R. K. (2020). Process mining in healthcare systems: a critical review and its future. *International Journal of Emerging Trends in Engineering Research*, *8*(9).

ter Voert, T. T. (2024). *Process Mining in Cyber Forensics - a Methodology for Process Mining in Forensic Investigation of Web Application Activity* – Tilburg University].

Tsai, C.-H., Liou, D.-K., & Lee, H.-L. (2024). Blockchain-supported online banking scheme. *Egyptian Informatics Journal*, *27*, 100516. https://doi.org/https://doi.org/10.1016/j.eij.2024.100516

van der Aalst, W., Adriansyah et al. (2012). Process Mining Manifesto. In F. Daniel, K. Barkaoui, & S. Dustdar, *Business Process Management Workshops* Berlin, Heidelberg.

van der Aalst, W. M. P. (2022). Process Mining: A 360 Degree Overview. In W. M. P. van der Aalst & J. Carmona (Eds.), *Process Mining Handbook* (pp. 3-34). Springer International Publishing. https://doi.org/10.1007/978-3-031-08848-3_1

van der Aalst, W. M. P., Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M., & Verbeek, H. M. W. (2007). Business process mining: An industrial application. *Information Systems*, *32*(5), 713-732. https://doi.org/https://doi.org/10.1016/j.is.2006.05.003

van der Werf, J. M. E. M., van Dongen, B. F., Hurkens, C. A. J., & Serebrenik, A. (2008). Process Discovery Using Integer Linear Programming. In K. M. van Hee & R. Valk, *Applications and Theory of Petri Nets* Berlin, Heidelberg.

van Dongen, B. F., & Van der Aalst, W. M. (2005). A Meta Model for Process Mining Data. *EMOI-INTEROP*, *160*, 30.

van Zelst, S. J., Mannhardt, F., de Leoni, M., & Koschmider, A. (2021). Event abstraction in process mining: literature review and taxonomy. *Granular Computing*, *6*(3), 719-736. https://doi.org/10.1007/s41066-020-00226-2

Verbeek, H. M. W., Buijs, J. C. A. M., van Dongen, B. F., & van der Aalst, W. M. P. (2011). XES, XESame, and ProM 6. In P. Soffer & E. Proper, *Information Systems Evolution* Berlin, Heidelberg.

Vona, L. W. (2016). *Fraud data analytics methodology: The fraud scenario approach to uncovering fraud in core business systems*. John Wiley & Sons.

Weijters, A. J. M. M., & Ribeiro, J. T. S. (2011, 11-15 April 2011). Flexible Heuristics Miner (FHM). 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM),