Received: October 18, 2021. Revised: December 20, 2021.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Hybrid Process Mining and Fuzzy Relation Weighting for Process Violation Weighting in Fraud Detection

Solichul Huda[1]* **Guruh Fajar Shidik[1]** **Fauzi Adi Rafrastara[1]** **Mohd. Faizal Abdollah[2]**

[1]*Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia*
[2]*Fakulti Tekonologi Maklumat and Komunikasi, Universiti Teknikal Malaysia Melaka, Malaysia*
* Corresponding author's Email: solichul.huda@dsn.dinus.ac.id

Abstract: There are many companies that implement Enterprise Resource Planning (ERP) to control their business processes. The risk of this implementation is the fact that fraud incidents in business processes also increase. Previous studies have proposed hybrid process mining with Fuzzy ARL and process mining with Heuristic Algorithm to detect fraud, but detection errors still occur because these methods cannot identify middle violations. This paper analyzes event logs in depth to determine employee relation weights during their activities. The relation weights obtained by hybrid with process mining are proposed to detect fraud. The proposed method integrates relational weights, process mining, and Fuzzy Multi-Attribute Decision Making to detect fraud. The relational weight method is used to determine the weight of the relation between employees. Process mining is used to compare the recorded event logs with the Standard Operating System (SOP). Finally, Fuzzy Multi-Attribute Decision Making is used to detect fraud. Using the same public dataset, the experimental results show that the process mining with Heuristic miner method obtained an accuracy of 0.9275, while process mining with the fuzzy ARL method obtained an accuracy of 0.9425, and process mining with the Relation weight obtained an accuracy of 0.96. Therefore, process mining and the Relation weight can detect fraud with medium violations and reduce false negatives.

Keywords: Fraud detection, Process mining, Anomalies, Business process.

1. Introduction

Adjusting to dynamic business changes, companies around the world use enterprise resource planning (ERP) to control their business processes [1, 2]. Business processes that run continuously make the number of process logs continue to grow. This condition makes it difficult for companies to analyze process logs manually. Thus, a method is necessary to analyze the process quickly and accurately [3-6].

The standard of the business process is made into a Standard Operating Procedure (SOP) which is used to control business processes. This is an advantage of implementing ERP in their company [7, 8]. This SOP is implemented to identify process violations [9, 10]. These violations can occur due to system errors or various other attributes [11-13]. However, these violations can result in fraudulent behavior [11]. Currently, company revenues are decreasing due to fraud incidents.

Fraud causes companies to suffer losses of up to 5% annually and increases every year by almost 1% [5, 14, 15]. This fraud often occurs in both medium and large companies. As a result, companies experienced a decline in their income.

Studies regarding fraud detection methods have been developed in previous studies with data mining and process mining approaches. Data mining analyzes input to build models and patterns used to test the process being examined. Several data mining methods such as decision trees and others were implemented in previous studies [16-18] to identify fraud in cases. However, these methods have weaknesses if the data is in the form of business process control flow. Previous research used process mining and Heuristic miner [10]; process mining and Fuzzy Association Rule Learning (ARL) [11] to

 DOI: 10.22266/ijies2025.0630.62

detect fraud in business processes. In this study, several process mining methods such as conformance checking, flow analysis, and pattern analysis are used to investigate event logs in business processes.

In previous research [10, 11], low level of violations is detected as not fraud, whereas low level violations in activities carried out by employees who have a strong relation weight with other employees who behave fraudulently will potentially be fraud. In this study, a method is proposed to detect fraud that commits low violations, which is carried out by employees who have strong relations with other employees who behave fraudulently. The proposed method integrates process mining, relation weighting, and Fuzzy Multi-Attributes Decision Making (MADM) to detect violations in business process.

This paper is outlined in five section. Section I is introduction of this study. Section II elaborated previous studies which applied process mining method for fraud detection. Section III describes the proposed methodology used in this study. Moreover, the relation weight is also expalained in detail in this section. Section IV explaned evaluation desain, dataset and discussion of research results. Section V points the conclution of this study.

2. Related works

Fraud is defined as an illegal profit-making activity [19]. Fraud can occur due to three possibilities, i.e. pressure or coercion, opportunity, and rationalization [20][21][22]. The use of ERP in companies makes SOPs become internal controls. When detecting fraud in a business process, internal control can be used to attack possible fraud [10],[11] [23]. The SOP for a business process must include a standard business process model, time records, resources, authority and decision-making. The complete SOP will become a reference in identifying violations from ongoing process standards [24], [25]. To analyze violations in business processes, techniques in process mining can be used [26], [27], [28],[29],[30].

Process Based Fraud (PBF) is a fraud that occurs in business processes. In previous research, attributes and patterns were identified to describe PBF. Previous studies on PBF [10], [11], [13] have identified attributes and patterns to explain fraud. Eight types of attributes of Fraud or fraudulent behavior in business processes can be distinguished i.e Skip activity, Wrong throughput time, Wrong pattern, Wrong resources, Wrong decision, Wrong duty, Parallel activities, and Wrong activity distance.

To detect violations, five process mining analyses are run: conformance checking analysis, times-stamp analysis, resource compliance checking, process data-flow and Time Between activities. Conformance-checking analysis can be run manually or with the help of the ProM application. This analysis is essential to identify fraud in the form of skip and wrong patterns. The implementation of this conformance checking uses conformance checking in ProM which compares the ongoing activity with the standard business process. The use of this conformance analysis is to determine the similarities and differences between the activities in the event logs with the standard business process. In this case, the activity that is running is different from the standard business process and is suspected of being a violation. This form of difference reveals fraudulent behavior.

Times-stamp analysis is used to analyze the activity execution time compared to the standard business process time. The implementation of times-stamp results in shorter and longer process execution times compared to the standard business process time. Anomalies in the form of throughput time min and throughput time max are identified by this method.

Resource compliance checking is an analysis method that compares employees (originators/resources) who carry out activities with the standard process in the SOP. This analysis method produces resources that violate the SOP in the form of wrong resources and wrong duties.

The process of data flow will analyze the flow of ongoing activities compared to SOP. This analysis will compare the sequence of ongoing activities compared to SOP. This method will produce activities that in making decisions violate SOP in the form of wrong decisions.

Time between activities will analyze the running distance time between two activities. The implementation of this method uses the time between activities in ProM. This analysis will compare the time and distance between the two activities compared to SOP. This method will produce violations in the form of parallel activities and distance activities.

2.1 Case study

In this case study, the business process of credit application was investigated to detect fraud. Analysis of the credit application process was used to identify activities that deviate from standard business processes. SOPs and business rules have been analyzed to obtain attributes at various rates.

The credit application process begins with checking the completeness of credit documents. After everything has finished and been completed, the

filing officer is handed over to the head of the office. Next, the head of the office gives recommendations to the officer to analyze the credit documents. After receiving the recommendation, the officer analyzes the applicant in more depth. If it is clear, the officer will verify the data at the location of the loan collateral (for example, personal property used as collateral for the loan) or at the debtor's office. If not, the credit application is rejected.

After verifying the collateral, the officer estimates the credit ceiling under the condition of the collateral, the character of the applicant, and the rules for granting credit. Next, the head of credit analysis checks the credit ceiling validity documents. If approved, the document is submitted to the credit administration for the document to be rechecked. The head of credit administration sends credit files to the head of the office by his authority. If the credit limit is approved by employees who are not authorized (wrong resource), then the result is a wrong decision. Next, the head of the office conveys the credit approval back to the credit administration section, who then submits the credit documents to the notary for the credit agreement process. If credit is rejected, the officer sends a rejection letter to the applicant. After the credit agreement is completed, the head of credit administration makes a withdrawal letter (transfer letter) and transfers the credit money to the applicant's account.

Credit application event logs analysis is used to analyze the credit application business process to obtain the sequence of activities, implementation time, employees who carry it out, separation of duties, and rules. If an activity is skipped, then the skipped attribute can be identified. If a case has an execution time activity that is longer than the standard execution time, then the throughput time attribute is flagged. Likewise, if the applicant's information is checked by an unauthorized author, the wrong resource attribute will be flagged. If the decision-making is not following the SOP, a wrong decision is made. When two or more activities are identified as being carried out by the same employee, the wrong duty is filled in. Then, if two activities are sequential but run simultaneously, the parallel activities of attributes will be filled in. Correspondingly, the time between running activities too quickly will be marked as activity distance. Overall, every SOP violation is associated with a Fraud attribute.

In the training session, the proposed method is carried out in three stages, i.e. conformance checking, relation weighting, and Fuzzy MADM. Conformance checking is part of process mining which is used to analyze business process violations of SOPs and employees on duty. In the study, the conformance checking methods used consist of skipped analysis, throughput time analysis, wrong pattern analysis, wrong resource analysis, wrong decision analysis, wrong duty analysis, wrong parallel activity analysis, and wrong activity distance analysis. Relation weighting is implemented to determine the weight of the relation between employees in carrying out activities. Fuzzy MADM is run to determine the violation rate. In this paper, the proposed conformance-checking methods include skip activity analysis, wrong throughput time analysis, wrong pattern analysis, wrong resource analysis, wrong decision analysis, wrong duty analysis, wrong parallel analysis, and wrong distance activity analysis. Fig. 1 shows the methodology proposed in this study.

3. Proposed methodology

This skip analysis uses the conformance checking plug-in in ProM that has been modified to identify skip activities. The inputs to this analysis are the series of activities in Petri net and event logs. If it is identified that an activity has been skipped and the activities are sequential, it will fill the skipped sequence attribute for that activity. Meanwhile, if the activity has branches, it will fill the skip decision attribute. This analysis generates the skip decision and skip sequence attributes.

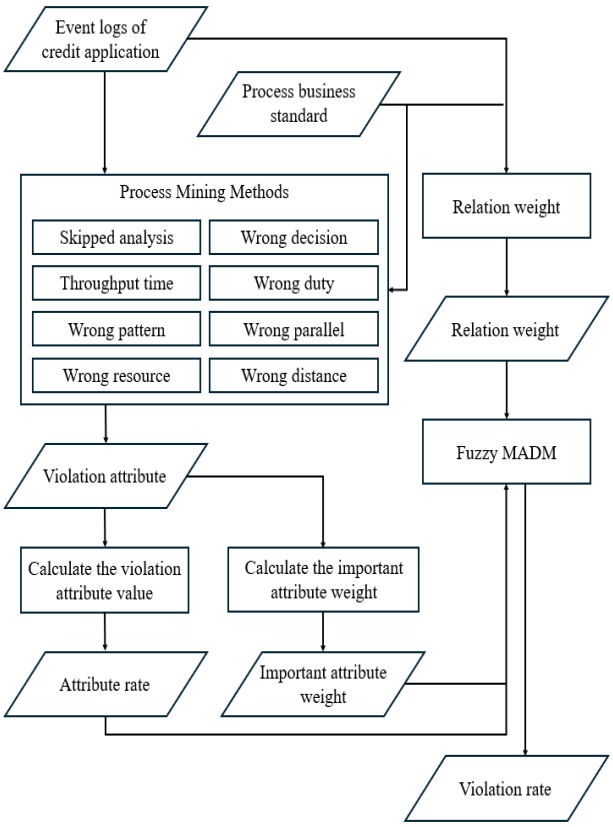

Figure. 1 The proposed method

3.1 Skipped activity analysis

This method analyzes the running time of an activity compared to the business process model. This analysis uses the Times-stamp analysis plug-in ProM which has been modified to analyze the activity running time shorter and longer than the standard time. If the activity running time is faster than the standard time, it will enter the throughput time min attribute. However, if the time of running the activity is slower than the standard time, it will add the throughput time max attribute. This method also produces the maximum time for the throughput time min and throughput time max attributes for each activity.

3.2 Wrong throughput time analysis

This method is used to analyze patterns in the course of a case. This analysis uses a modified conformance-checking plug-in in ProM to identify patterns of ongoing activity compared to patterns of activity in the standard model. The input of this analysis consists of a series of activities in the form of Petri nets and event logs. This analysis produces wrong pattern attributes.

3.3 Wrong pattern analysis

This method will analyze the authority of employees running the activity and the authority of employees in the SOP. This analysis uses a modified resource compliance checking plug-in in ProM to compare with the SOP. If the authority of the employee carrying out the activity is not under the SOP, a wrong resource violation is identified. This analysis generates the wrong resource attribute.

3.4 Wrong resource analysis

This method is used to analyze the determination of activity flows that are not based on the SOP. This analysis is run using the process data-flow plug-in in ProM that has been developed to identify decision-making errors. In business processes, some activities determine one option. If the selected option does not comply with the SOP, it will result in a wrong decision.

3.5 Wrong decision analysis

Wrong duty is one of the methods used to analyze the occurrence of duplicate duties. The resource compliance checking plug-in in ProM that has been developed is used to analyze the assignment of activities for each employee. This analysis produces the wrong duty attribute.

3.6 Wrong duty analysis

In the SOP, some activities are executed sequentially or simultaneously. There are times when employees carry out two activities simultaneously, even though according to the SOP they should be carried out sequentially, these processes are identified as parallel activities. This analysis is implemented using the time between activities plug-in in ProM which has been modified to analyze activities executed in parallel or sequentially. This analysis generates the parallel activities attribute.

3.7 Wrong parallel activity analysis

In the SOP, some activities are executed sequentially or simultaneously. There are times when employees carry out two activities simultaneously, even though according to the SOP they should be carried out sequentially, these processes are identified as parallel activities. This analysis is implemented using the time between activities plug-in in ProM which has been modified to analyze activities executed in parallel or sequentially. This analysis generates the parallel activities attribute.

3.8 Wrong activity distance analysis

This method will analyze the time distance between two activities, i.e. the end time of the activity and the start time of the next activity. This analysis also uses the modified time between activities plug-in in ProM to analyze the time distance between an activity and the next activity and compare it with the SOP. This analysis produces the wrong activity distance attribute.

All violations of SOP obtained are trained using relation weight and fuzzy multi-attribute decision-making. The method consists of two ways. First, calculate the weight of the relations between employees. This process is done using the relation weighting method proposed by this study. Second, calculate the violation rate for each case. This process is done in two ways. First, determine the important attribute weight according to expert opinion. Second, determine the violation attribute rate. Then the violation attribute rate is adjusted to the relation weight of employees who carry out the activity. Then, the attribute rate and relation weight are adjusted to the important attribute weight. Finally, determine the violation rate. The input for this process is the violation rate of all identified attributes. This process is carried out using fuzzy multi-attribute decision-making.

3.9 Relational weighting

A case in event logs is a series of business processes starting from the first activity to the last activity. Each activity is done by an employee (originator). Each activity that is run must be preceded by an activity and followed by the theme next activity. Likewise, the employee who runs the activity is preceded by the employee who runs the previous activity and the employee who runs the activity after. The frequency of the sequence between employees in running this activity is the basis for calculating the relational weight between employees which in this paper is known as relation weighting. This relation weighting is the contribution of this paper in detecting Fraud.

The number of activities from the case being run is increasing, making the number of relations between employees increase. The magnitude of relations between employees who are in direct sequence in carrying out activities affects the weight of the relations between them. The number of direct relations is proposed to explore the weight of closeness between employees. The weight of relations between employees is introduced as the weight of relations in detecting fraud, which is the main contribution of this paper.

The sequence of activities in a case shows the sequence of employees who carry out activities in a case. Based on conformance checking, employees will be identified as carrying out activities according to or violating SOP. Furthermore, in training, the identified violations are given to experts to be assessed as fraud or ordinary violations. In this paper, an employee who has a strong relation with another employee who is determined as a perpetrator of fraud, then the weight of the relation becomes a variable for calculating the weight of the violation.

This study explores new knowledge from event logs to analyze minor violations in more detail. The weight of direct relations between employees is measured from the number of direct relations between employees in carrying out the activities of a case. Then, based on the number of relations, the probability value is calculated between the number of direct relations compared to the total direct relations that occur in the event logs. The probability value is used as the weight of the relations between employees. In addition to the relation weight, the method proposed in this study is different from previous studies [10], [11]. In this paper, the calculation of the rate attribute is done at the activity level. For example, the occurrence of throughput time max in the Get_Info activity. To determine the throughput time max in the Get_info activity from the training data, it is found to be 10 minutes, and then the membership class is determined based on the value of 10. Then, to determine the rate attribute throughput time max is determined based on all throughput time max events that occur in the case. This has not been proposed by previous studies. Eq. 1 is the relation weight between the two employees proposed by this paper.

$$RM = LE_1 \rhd LE_2 = \frac{(\sum_{c \in L} |E_1 \rhd E_2|)}{(\sum_{c \in L} |c|-1)} \quad (1)$$

where RM is the relation weight, E_1 is the 1st employee, E_2 is the next employee, C is the case and L is the transaction logs. $\boldsymbol{LE_1 \rhd LE_2}$, this function will return the true if E_1 and E_2 run the same case activities and the distance between these two activities is one. In the credit application log, a case contains a set of credit application business processes. Each case begins with the activity of receiving credit application documents and ends with the activity of credit rejection or credit disbursement. Each activity is run by an employee. This way, every time there is a violation, it will be known to the employee carrying it out. Correspondingly, employees who run an activity know the employee who will run the following task.

For example, an activity of check completeness is carried out by James, and the next activity is check SID which is run by Olsen. In the event logs, it turns out that Jeff and Olsen worked on these activities sequentially 560 times. Thus, they have interacted 560 times. An activity of check collateral documents is carried out by Devan, and the next activity is check collateral type which is run by Jack. In the event logs, it turns out that Delano and Jack worked on these activities sequentially 980 times.

Initial research shows that the intensity of interaction between employees in carrying out business processes will influence the weight of the relation between the two. Fig. 2 shows an example of two employees interacting directly.

This credit application consists of 23 activities run by 23 different employees. In one case, relations between employees occur at different distances. The gap between employees is determined by the number of activities between them. The further the distance between the two activities they are working on, the smaller the weight of the interaction. For example, there are three activities, i.e. the activity of check completeness is performed by Jeff, the second activity check SID is carried out by Olsen, and the third activity checking collateral documents is run by Devan. In this case, Jeff and Olsen are one activity apart, while Jeff and Devan are two activities apart.

This difference affects the weight of the relation between them. In this study, what is calculated is the weight of relation between employees who interact directly or at a distance.

In the training, 23 employees were identified for each case who interacted directly. As a result, the relation weight method identifies employees who interact directly. The following Table 1 shows the number of employees who interact directly.

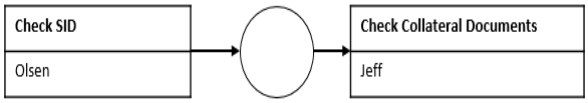

Figure. 2 An example of two employees interacting directly

Table 1. Example of the number of consecutive employees.

Employees 1	Employee 2	Direct interaction
Jeff	Olsen	560
Yaron	Wesly	451
Harry	Dion	900
Patrick	John	890
Shane	Brian	1000
Devan	Jackob	1000
Yaron	Olsen	880
Harry	John	910
Delano	Jack	980
Sean	Patrick	451

Table 2. Level of relation weight

	Fuzzy Parameters				Scale
Levels	A	B	C	D	
Strong	0.3	0.4	1	1	100% - 30%
Fair	0.1	0.2	0.3	0.4	40% - 10%
Weak	0	0	0.1	0.3	0% - 30%

Table 3. Example of the weight of relation between employees

Employees 1	Employee 2	Direct interaction
Jeff	Olsen	0.0205
Yaron	Wesly	0.0205
Harry	Dion	0.040909
Patrick	John	0.040455
Shane	Brian	0.040455
Devan	Jackob	0.040455
Yaron	Olsen	0.04
Harry	John	0.041364
Delano	Jack	0.044545
Sean	Patrick	0.0205

Table 4. Example of fuzzy relation weights.

Employees 1	Employee 2	Direct interaction
Jeff	Olsen	Fair
Yaron	Wesly	Fair
Harry	Dion	Strong
Patrick	John	Strong
Shane	Brian	Strong
Devan	Jackob	Strong
Yaron	Olsen	Strong
Harry	John	Strong
Delano	Jack	Strong
Sean	Patrick	Fair

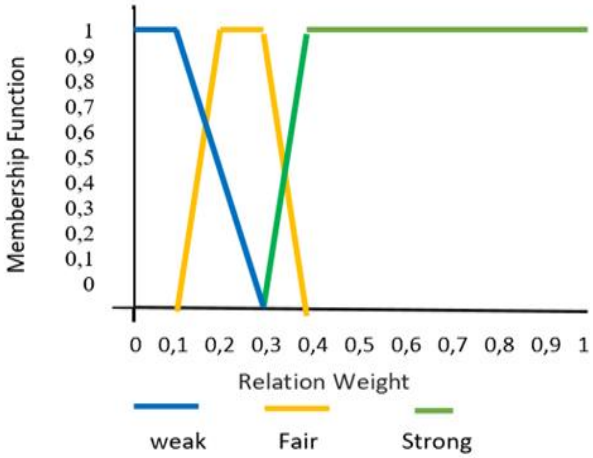

Figure. 3 Relation weight membership function

These relation weights between 0 – 1 are determined in three categories, namely weak, fair, and strong. The relation weight levels are shown in Table 2. Meanwhile, the relation weight membership function is shown in Fig. 3. Examples of relation weights and relation weights levels from training data are respectively illustrated in Table 3 and Table 4.

This study analyzes the business process of credit application. The methods of skipped analysis, throughput time analysis, wrong resource analysis, wrong duty analysis, wrong pattern analysis, wrong decision analysis, wrong activities parallel analysis and wrong distance activity analysis were employed to analyze the occurrence of SOP violations. Then, the weight of the violations identified was calculated.

During the training stage, activities were also identified that violated the SOP and the employees who carried them out. This process produces employees who commit violations along with the type and name of the activity. The results show the number of violations committed by employees and the types of violations in credit applications. Examples of employees' names and types of violations committed are described in Table 5.

Table 5. Examples of employee violations

No.	Employee Name	Type of Violation	Num. of Cases
1	Jeff	Throughput time Min, wrong duty	5
2	Yaron	Throughput time max	1
3	Olsen	Skip	6
4	Devan	Wrong pattern	1
6	Brian	Throughput time min, throughput time max	2

Table 6. Examples of expert assessments of employee violations

No.	Employee Name	Type of Violation	Category
1	Jeff	Throughput time Min, wrong duty	Fraud
2	Yaron	Throughput time max	No
3	Olsen	Skip	Fraud
4	Devan	Wrong pattern	No
6	Brian	Throughput time min, throughput time max	No

Moreover, the expert, based on his experience, provides an assessment of whether the employee is a fraud perpetrator or not. This category is a consideration of the weight of the relation and determines the severity of the violation or not. If an employee has a relation with an employee who is involved in fraud, then the weight of the relation is used in determining the severity of the violation. However, if it is not fraud, then the relation weights are not used. Table 6 shows examples of the types of violations for each employee and their categorization by experts.

3.10 Fuzzy multi-attribute decision making

This method is used to determine the violation rate of violation activities in a case. Determining the violation rate implements two concepts, namely fuzzification and multi-attribute decision-making (MADM). MADM can be used to determine the choice of several alternative values. However, MADM has the weakness of being less accurate for alternative values in the form of linguistic information. Therefore, fuzzification is needed to handle this linguistic information.

Three data are needed to determine the level of violation, namely the occurrence of violations obtained from conformance checking, the assessment of the importance weight of the violation attributes by experts, and the weight of the relation between employees during the activity. The three data are converted into fuzzy numbers based on the level and membership function.

Table 7. Violation units for each attribute

No.	Event name	Unit
1.	Throughput time min	Minutes
2.	Throughput time max	Minutes
3	Wrong pattern	Activities (Number of different events)
4	Wrong decision	Activities (Number of options available)
5.	Wrong resources	Level (Level skipped)
6.	Wrong duty	Events (Number of events run)
7.	Parallel events	Events (Number of events running simultaneously)
8.	Distance events	Minutes

Table 8. Maximum violations for each activity

No.	Attribute Name	Activity name	Max. Amount of TT. Min	Max. Amount of TT.Max
1	Throughput time	Receive_ application	10	30
2	Throughput time	Check_ completeness	5	10
3	Throughput time	Check_ SID	5	5
4	Throughput time	Check_ collateral Document	20	30
5	Throughput time	Check loan type	2	10
6	Throughput time	Collateral_ verification _locate	180	180
7	Throughput time	Collateral_ local _government	60	120
8	Throughput time	Collateral_ government	60	120
9	Throughput time	Complete_ verification	30	60
10	Throughput time	Ceiling_ estimation	20	30

The phase of determining violations is carried out with determined the attribute value. The input of this process is the violation value and the maximum violation value in the activity. The violation value is obtained from the conformance method. While the maximum violation value is obtained from the highest violation value in the activity. For example, running the collateral check activity for 25 minutes, while the standard runs the activity for 10 minutes.

Table 9. Linguistic accuracy

Levels	Fuzzy Parameter			
	A	B	C	D
Very strong	0.8	0.9	1	1
Between very strong and strong	0.7	0.8	0.9	1
Strong	0.5	0.6	0.7	0.8
Between fair and strong	0.4	0.5	0.6	0.7
Fair	0.3	0.4	0.5	0.7
Between fair and weak	0.2	0.3	0.4	0.5
Weak	0.1	0.2	0.3	0.4
Between weak and very weak	0	0.1	0.2	0.3
Very weak	0	0	0.1	0.2

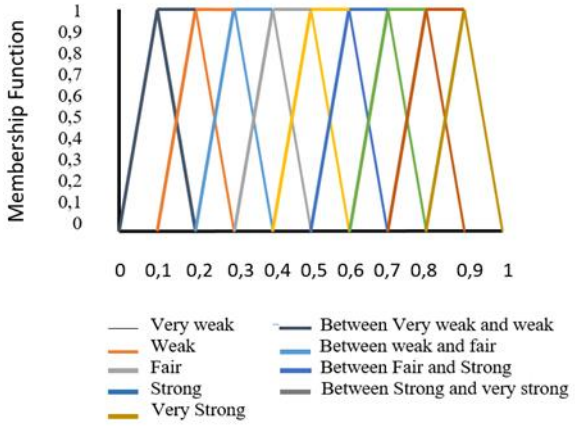

Figure. 4 Membership function of attribute value

Thus, the throughput time max attribute occurs at this activity. Furthermore, we calculated the attribute value. The input to this process is the number of violations and the maximum value which occurs during the activity. For example, if the maximum throughput time value for the activity is 16, then the violation weight is 15/16. Eqs. (2) and (3) each are used to calculate the maximum violation for each activity and calculate the attribute value. Table 7 and Table 8 show each example of an attribute unit and the maximum violations for each attribute for each activity.

$$Maximum\ Violation = Max\ (Attribute_i) \quad (2)$$

$$Av = \frac{violation\ value\ of\ an\ attribute}{Max\ viol\ of\ an\ att\ every\ activity} x100\% \quad (3)$$

The attribute value obtained, for the process of processing violation data, the attribute value is converted into the violation level presented in Table 9. For the membership function, the attribute value uses the membership function as depicted in Fig. 4. The range from 0 to 1 attribute value is divided into 9 categories to determine the parameters of the membership functions A, B, C, and D.

The membership function formula can be found in Eqs. (4) to (7), and the membership function parameters are explained in Table 10.

$$Bottom = \frac{\sum_{k=1}^{j} a_k}{j} = \frac{0+0+0+0}{4} \quad (4)$$

$$Middle\ bot = \frac{\sum_{k=1}^{j} a_k}{j} = \frac{0{,}2+0{,}2+0{,}2+0{,}2}{4} \quad (5)$$

$$Middle\ Top = \frac{\sum_{k=1}^{j} a_k}{j} = \frac{0{,}3+0{,}3+0{,}3+0{,}3}{4} \quad (6)$$

$$top = \frac{\sum_{k=1}^{j} a_k}{j} = \frac{0{,}4+0{,}4+0{,}4+0{,}4}{4} \quad (7)$$

In this study, the Expert assessment used is based on the expertise of a bank's credit application auditor. Table 11 shows the results of the assessment of four experts on the violation attributes. For the membership function, the importance weight attribute uses the membership function as depicted in Fig. 5.

Table 10. Membership function parameter of attribute value

Membership function of VW		Membership function of BVW and W, W, BW and F, F, BF and S, S, BS and VS		Membership function of VS	
Degree	Condition	Degree	Condition	Degree	Condition
1	a≤x≤C	0	x≤a	0	x≤a
(d-x)/(d-c)	c < x < d	(x-a)/(b-a)	a<x<b	(x-a)/(b-a)	a < x < b
0	x ≥d|x<a	1	b≤x≤c	1	x ≥ b
		(d-x)/(d-c)	c<x<d		
		0	x≥d		

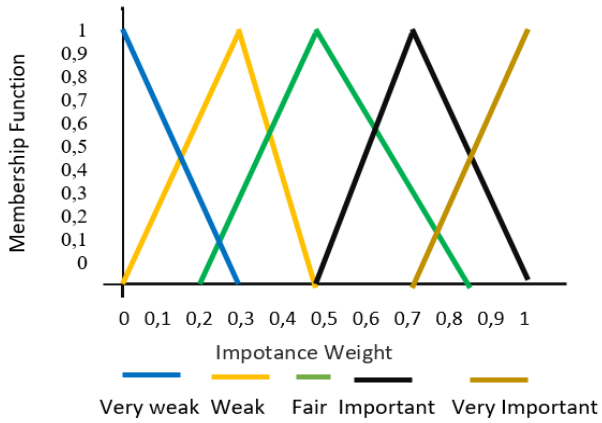

Figure. 5 Membership function of importance weight attribute.

Table 11. Assessment of attribute weights by experts

Anomaly Attributes	Expert 1	Expert 2	Expert 3	Expert 4
Skip sequence	VI	VI	I	VI
Skip decision	VI	VI	VI	VI
Throughput time min	W	W	W	W
Throughput time max	W	W	F	W
Wrong resource	I	I	I	I
Wrong decision	VI	VI	VI	VI
Wrong duty	W	W	W	W
Parallel event	F	F	F	F
Wrong event distance	I	I	I	I

Table 12. Level of importance weight attribute

Levels	Fuzzy parameters				Scale
	A	B	C	D	
Very Important	0.7	1	1	1	100%-70%
Important	0.5	0.7	0.7	1	100%-50%
Fair	0.2	0.5	0.5	0.8	80%-20%
Weak	0.1	0.3	0.3	0.5	50%-10%
Very Weak	0	0	0	0.3	0%-30%

A weight measurement is carried out for each violation attribute based on Table 11. The value weight is divided into four categories, namely Bottom, middle bottom, middle top, and top. The calculation of the four category weights is carried out by implementing formulas 3, 4, 5 and 6, where j is the number of experts, and the contents of the variables A, B, C, and D are the values of the vectors A, B, C, D in Table 12. In Table 12, the range from 0 to 1 is divided into five categories to determine the parameters of the membership functions A, B, C, and D. The results are shown in Table 13.

The probability of violation is measured for each violation attribute. In calculating the occurrence rate of each violation attribute, three variables are used, namely the conformance result value of the examination, the importance weight given by the experts, and the relations weight between employees. The adjustment of violations (Attribute value) with the importance weight of the expert assessment attribute is carried out using Eq. (7). The values of AV_1, AV_2, AV_3, and AV_4 in Eq. (7) are the vector values A, B, C, and D of the fuzzification value of the violation as in Table 13. In addition, the values of FW_1, FW_2, FW_3, and FW_4 in Eq. 7 are the vector values A, B, C, and D in Table 12. The RA value reflects the fuzzification into the violation class based on the membership function of the occurrence level of the violation attribute. For example, the conformity examination result for the violation is Between Very Weak and Weak, with the first expert assessment for the weak category. Therefore, the adjustment value is given by Eq. 7. Then the RA value = 0.2125 is obtained reflecting the fuzzification into the violation class based on the weak membership function. The adjustment results of this example can be seen in Table 14.

Table 13. Weight of attribute violations by expert.

Anomaly attributes	Bottom	Mid Bot	Mid Top	Top
Skip sequence	0.65	0.925	0.925	0.925
Skip decision	0.7	1	1	1
Throughput time min	0.1	0.3	0.3	0.5
Throughput time max	0.125	0.35	0.35	0.575
Wrong resource	0.5	0.7	0.7	1
Wrong decision	0.7	1	1	1
Wrong duty	0.1	0.3	0.3	0.5
Parallel event	0.2	0.5	0.5	0.8
Wrong event distance	0.5	0.7	0.7	1

Table 14. Assessment of attribute weights by experts

Anomaly Attributes	Exp 1	Exp 2	Exp 3	Exp 4	Linguis-tic
Skip sequence	0	0	0	0	0
Skip decision	0	0	0	0	0
Throughput time min	W	W	W	W	W
Throughput time max	0	0	0	0	0
Wrong pattern	W	W	W	W	W
Wrong resource	0	0	0	0	0
Wrong decision	0	0	0	0	0
Wrong duty	0	0	0	0	0
Parallel event	0	0	0	0	0
Wrong event distance	0	0	0	0	0

Table 15. Assessment of attribute weights by experts

Anomaly Attributes	Bottom	Mid Bot	Mid Top	Top	Linguis-tic
Skip sequence	0	0	0	0	0
Skip decision	0	0	0	0	0
Throughput time min	0.05	0.15	0.2	0.35	0.1875
Throughput time max	0	0	0	0	0
Wrong pattern	0.4	0.05	0.6	0.85	0.6
Wrong resource	0	0	0	0	0
Wrong decision	0	0	0	0	0
Wrong duty	0	0	0	0	0
Parallel event	0	0	0	0	0
Wrong event distance	0	0	0	0	0

After being adjusted to the attribute importance weight, the calculation results are adjusted to the relation weight between employees who are directly related. The adjustment uses Eq. 8. In this calculation, the WA_1, WA_2, WA_3, and WA_4 values in Eq. 8 are the vector values A, B, C, and D in Table 14 based on the fuzzification value of the violation value adjustment with the attribute importance weight. In addition, in this calculation, the SN_1, SN_2, SN_3, and SN_4 values are the vector values A, B, C, and D in Table 2 (employee relation weight). The RR value reflects the fuzzification into the violation class (attribute value) and the attribute importance weight with the relation weight, based on the violation membership function, the attribute importance weight, and the relation weight. For example, the results of the conformity check for violations and attribute importance weights are in Between weak, with the relation weight for the strong category. Therefore, the adjustment value is given by Eq. 8. Then the RR value = 0.4125 is obtained reflecting the fuzzification into the violation class (attribute value) based on the Fair membership function (Eq. (9)).

$$RA = \frac{\frac{(AV_1 + AV_2 + AV_3 + AV_4)}{4} + \frac{(FW_1 + FW_2 + FW_3 + FW_4)}{4}}{2} \quad (8)$$

$$RR = \frac{\frac{(WA_1 + WA_2 + WA_3 + WA_4)}{4} + \frac{(SN_1 + SN_2 + SN_3 + SN_4)}{4}}{2} \quad (9)$$

After obtaining the attribute rate for each identified activity, the rate attribute of all attributes in a case is determined (Table 15). The input to this process is the attribute rate for each activity. The result is a violation value for each attribute identified in a case. For example, if the attribute throughput time max is identified for three activities, then the attribute rate for the three attributes is calculated. Eq. (10) is used to obtain the rate attribute for each of the same attributes identified in a case.

$$\mathrm{WA_n} = (WA_1 \; V \; WA_2 \; V \; WA_3 \; ... \; V \; WA_n) \quad (10)$$

Where WA is the attribute rate for each activity, n is the number of similar attributes in a case.

In the context of fraud detection, the rate of the violations identified in a case must be determined. In this paper, the weight of the violation is defined as the fraud rating. The input to this process is the attribute rate of all identified attributes. For example, in a case four attributes are identified, including wrong throughput time min, wrong duty, wrong resource, and wrong activity parallel, then the fraud rating is calculated based on these four attributes. Eq. (11) is used to determine the fraud rating.

$$Fr = \frac{1}{k}[(WF_1 + WF_2 + WF_3 \; ... + WF_k)] \quad (11)$$

Where WF is the weight attribute, k = number of attributes in a case.

3.11 Fraud threshold

This research was conducted to increase the accuracy of the proposed fraud detection method compared to fraud detection using process mining with Heuristic miner [10]; and using process mining with Fuzzy Association Rule Learning (ARL) [11]. Therefore, this study determines the same threshold limit as research [11] i.e. 30% for fraud. The similarity of the threshold values will make the advantages of the proposed method clear.

4. Result and discussion

4.1 Evaluation design

The evaluation in this research focuses on the following: (1) finding the advantages of process minng with the relation weight in determining the rate of violation compared to using process mining with the Heuristic miner method [10]; and process mining with the fuzzy ARL method [11] in the context of fraud detection, then (2) measuring the accuracy of third methods. The scenarios and datasets used in this evaluation are the same for both methods. This experiment is based on a case study of credit applications at banks. The dataset consists of a training dataset and a testing dataset generated by several distribution models as provided in [31].

Based on the analysis, violations are modeled as attributes using a Poisson distribution with parameters set to 3. This parameter shows that on average, there are 3 unusual cases every month. The Poisson distribution is used because its characteristics are in line with business process fraud behavior. The overwhelming number of cases for each attribute is generated randomly based on a Poisson distribution. Therefore, each attribute has a different number of extraordinary cases each month.

Table 16. Example of violation attribute for each case using poisson distribution

Case Number	Activities name						
	Receive applications	Check completeness	Give info	Check SID	Check collateral_document	Check loan_type	Collateral verify_locate
601	-	-	-	-	-	-	Throughput time min
602	-	-	-	-	-	-	-
603	-	-	-	-	-	-	-
604	-	-	-	-	-	-	-
605	-	-	-	-	-	-	-
606	-	Throughput time min	-	-	Throughput time min	-	-
607	-	-	-	-	-	-	-
608	-	-	-	-	-	-	-
609	-	-	-	-	-	Wrong resource	-
610	-	-	-	-	-	-	-

Table 17. Example of the number of violations for each attribute using discrete distribution

Case Number	Skip		Wrong throughput time		Wrong pattern	Wrong resource	Wrong decision	Wrong duty	Wrong parallel event	Wrong event distance
	Sequence	Decision	Min	Max						
601	0	0	1	0	0	0	0	0	0	0
602	0	0	0	0	0	0	0	0	0	0
603	0	0	0	0	0	0	0	0	0	0
604	0	0	0	0	0	0	0	0	0	0
605	0	0	0	0	0	0	0	0	0	0
606	0	0	2	0	0	0	0	1	0	0
607	0	0	0	0	0	0	0	0	0	0
608	0	0	0	0	0	0	0	0	0	1
609	0	0	0	0	0	1	0	0	0	0
610	0	0	0	0	0	0	0	0	0	0

Table 18. Fraud Determination Role

No.	Fraud	Non-Fraud
1	Cases that contain skips	Cases without attributes are skipped
2	Cases containing wrong decisions	Cases without the wrong decision attribute
3	Cases containing wrong resources and wrong duties in one case	Cases that contain either wrong resources or wrong duties
4	Cases that have attributes	Case without any attributes
5	Cases that contain more than one attribute	Cases that have only 1 attribute
6	Cases that obtain a fraud rate equal to or greater than 0.3	Cases that have a fraud rate of less than 0.3

Table 19. Comparison of Accuracy Test Results for Heuristic Miner, Fuzzy ARL, and the Proposed Relation Weight Method

Method used		True Class	
		Positive	Negative
Heuristic miner [10]	Positive	161	10
	Negative	19	210
Fuzzy ARL [11]	Positive	167	10
	Negative	13	210
Relation Weight (*Proposed Method*)	Positive	174	10
	Negative	6	210

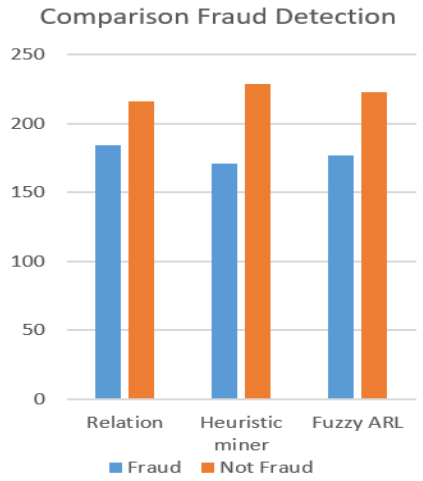

Figure. 6 Comparison of fraud detection methods

In addition, 50 credit applications are processed every month. The violations are spread among all credit applications based on a uniform (discrete) distribution. The goal is to randomly distribute violations across more than 50 credit applications a month, based on the number of violation activities for each attribute. Examples of the resulting data can be seen in Tables 16 and 17.

Overall, 2000 cases were generated as experimental data. The experimental data were then divided into training data and test data. In the training data, there were 1600 cases consisting of 848 fraud cases and 752 legal cases. Meanwhile, in testing data, there were 400 cases consisting of 180 fraud cases and 220 legal cases.

To analyze process violations from event logs, we developed the ProM plug-in. This plug-in is used to analyze the conformity of business processes with SOPs. Then, the data in the event logs is trained with additional plug-ins using conformance checking (ProM is also developed according to training needs).

From the analysis conducted on the testing data, cases that violate and those that comply with the SOP are identified. Examples of violation attributes and details of violations for each case are shown in Tables 16 and 17 respectively. Furthermore, the violation unit and the maximum number of violations for each attribute are determined. Tables 17 and 8 each show the maximum violations along with their units and the maximum total violations that occur for each attribute.

To compare the weights of relations with Heuristic miner and fuzzy ARL, training with the weights of relation; training with Heuristic miner; and training with fuzzy ARL using the same data. In the training stage, expert assessments are used to determine whether a case is fraud or not. In this study, experts also provide a role for fuzzy ARL as in Table 18.

4.2 Discussion

To clarify the advantages of using process mining with the relation weighting method compared to process mining with Heuristic miner [10]; and process mining with fuzzy ARL [11], it is necessary to analyze several case examples of the three methods.

On the other hand, with their expertise, the experts analyzed cases in event logs. Fig. 7 illustrates a comparison of the three methods of detecting fraud.

In measuring the accuracy of these three methods, evaluation with receiver operating characteristic (ROC) framework analysis was performed by using Eq. (12). The results of accuracy measurements for process mining with the Heuristic miner method [10], process mining with the fuzzy ARL method [11], and process mining with the relation weight as a proposed method are shown in Table 19.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

From the results of accuracy measurements, it can be seen that the use of process mining with relation weight in weighting violations has higher accuracy than process mining with Heuristic miner [10]; and process mining with Fuzzy ARL [11]. Accuracy with relation weights reached 0.96, while Heuristic miner was valued at 0.9275, while Fuzzy ARL was valued at 0.9425. A comparison of the three methods of detecting fraud is shown in Fig. 6.

Accordingly, using relation weights between employees in investigating SOP violations obtains better accuracy. From testing, relation weight obtained better accuracy than Heuristic Miner [10] and Fuzzy ARL [11], because the proposed method can reduce false negatives.

This study was conducted by analyzing event logs from a credit application dataset, i.e. a public dataset. This dataset has been used by several previous studies including [11, 31]. In addition to detecting fraud in credit applications, this proposed method can also be used to detect fake policy submissions (Fraud) in event logs from insurance claim applications or Fraud in event logs from procurement applications or Fraud in event logs from health cost applications. This ability is due to the similarity of the event log types from the various applications.

5. Conclusion

Based on the experimental results, the relation weights obtained from event logs can describe the level of relations between employees while carrying out activities. In addition, it can be concluded that the process mining approach with the method of relation weight, fuzzy ARL and Heuristic miner can be applied to detect fraud in business processes. The process mining method can identify violations in business processes by checking the conformity between business processes and SOPs. The relation weight method; Heuristic miner and fuzzy ARL methods are trained using the same data to determine fraud in business processes. In analyzing testing data. Process mining and the heuristic miner obtained an accuracy of 0.9275, process mining and the fuzzy ARL method obtained an accuracy of 0.9425, while process mining and the relation weight obtained an accuracy of 0.96. This shows that the relation weight can detect fraud more accurately at middle violations. In addition, the combination of process mining and relation weighting can identify fraud with middle violations. Hence, process mining and the relation weighting methods can be used to help identify fraud in cases of minor violations, so that fraud incidents due to employee relations are easily detected.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

This work is a contribution of the authors: Conceptualization, Solichul Huda and Guruh Fajar Shidik; methodology, Solichul Huda; validation, Solichul Huda and Mohd. Faizal Abdollah; formal analysis, Solichul Huda and Guruh Fajar Shidik; writing—original draft preparation, Solichul Huda. writing—review and editing, Mohd. Faizal Abdollah and Fauzi Adi Rafrastara.

References

[1] M. C. Abejo, “Enterprise Resource Planning System Implementation Framework for Selected State Universities”, *International Journal of Computing Sciences Research*, Vol. 7, pp. 2450-2477, 2023.

[2] V. Narayan1. S. Ganapa, “ Hybrid Sampling and Similarity Attention Layer in Bidirectional Long Short Term Memory in Credit Card Fraud Detection”, Vol. 15. No. 6, PP. 35-244, 2022.

[3] D. Trisanto, N. Rismawati, M. F. Mulya F. I. Kurniadi, “ Effectiveness Undersampling Method and Feature Reduction in Credit Card Fraud Detection”, Vol.13, No.2, pp. 173-181 2020

[4] T.S. Wulan, P.W. Novika, E. Novianti, and F.A. Putra, “Impact of ERP System Implementation on Operational and Financial Efficiency in Manufacturing Industry”, *Journal of Economic Education and Entrepreneurship Studies*, Vol. 5, No. 3, 2024

[5] M. Riskiyadi, “Detecting future financial statement fraud using a machine learning model in Indonesia: a comparative study”, *Asian Review of Accounting*, Vol. 33, No. 1, 2023.

[6] S. Bernardi, R. P. Alastuey, and R. Trillo-Lado, “Using Process Mining and Model-Driven Engineering to Enhance Security of Web Information Systems”, *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 160-166, 2017, doi: 10.1109/EuroSPW.2017.66.

[7] W. M. P. Van Der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 2011, doi: 10.1007/978-3-642-19345-3.

[8] G. Zioviris, K. Kolomvatsos, G. Stamoulis. “ An intelligent sequential fraud detection model based on deep learning”, *The Journal of*

Supercomputing, Vol. 80, pp. 14824-14847, 2024

[9] S. Huda, R. SarNo. and T. Ahmad, "Fuzzy MADM Approach for Rating of Process-Based Fraud", *J. ICT Res. Appl.*, Vol. 9, No. 2, pp. 111-128, 2015, doi: 10.5614/itbj.ict.res.appl.2015.9.2.1.

[10] M.C.d. Silva, G. M. Tavares, M.C. Gritti, P.C,Ceravolo, and S.B. Junior, "Using Process Mining to Reduce Fraud in Digital Onboarding", *FinTech Journal*, 2, 120-137, 2023.

[11] R. SarNo. F. Sinaga, and K. R. SungkoNo. "Anomaly detection in business processes using process mining and fuzzy association rule learning", *J. Big Data*, Vol. 7, No. 1, p. 5, 2020, doi: 10.1186/s40537-019-0277-1.

12] R. Sarno and K. R. SungkoNo. "A survey of graph-based algorithms for discovering business processes", *Int. J. Adv. Intell. Inform.*, Vol. 5, No. 2, p. 137, 2019, doi: 10.26555/ijain.v5i2.296.

[13] S. Huda, Aripin, M. F. Naufal, V. Martianova Yudianingtias, and Anisti, "Fraud Patterns Classification: A study of Fraud in business Process of Indonesian Online Sales Transaction", In: *Proc. of 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)*, pp. 212-217, 2020, doi: 10.1109/MECnIT48290.2020.9166644.

[14] ACFE, *Occupational Fraud 2024:A Report To The Nations*®. [Online]. Available: https://legacy.acfe.com/report-to-the-nations/2024/

[15] K. R. Prasanna Kumar, S. Aravind, K. Gopinath, P. Navienkumar, K. Logeswaran & M. Gunasekar, " Enhancing the Credit Card Fraud Detection Using Decision Tree and Adaptive Boosting Techniques", In: *Proc. of 22nd International Conference on Intelligent Systems Design and Applications (ISDA 2022)*, pp. 358-365, doi: https://doi.org/10.1007/978-3-031-35501-1_36.

[16] Z. Hamid, F.Khalique, S. Mahmood, A. Daud, A. Bukhari and B. Alshemaimri, ", *Healthcare insurance fraud detection using data mining*, Vol 24:112., 2024.

[17] A. Raslan, S.Ali", Using Data Mining Techniques for Fraud Detection in The Non-banking Sector", *Journal of Computing and Communication*, Vol.3 , No.1 , pp. 132-142, 2024

[18] T. Malekolkalami, K.K. Parijan, M. Alifari, "Application of Data Mining to Detect Accounting Fraud in Information Systems", *International Journal of Knowledge Processing Studies*, Vol. 3, No. 4, pp. 60-72, 2023.

[19] S. Huda, A. -, M. F. Naufal, and V. M. Yudianingtias, "IDENTIFICATION OF FRAUD ATTRIBUTES FOR DETECTING FRAUD BASED ONLINE SALES TRANSACTION", *Indian J. Comput. Sci. Eng.*, Vol. 12, No. 5, pp. 1409-1424, 2021, doi: 10.21817/indjcse/2021/v12i5/211205083.

[20] V. Narayan S. Ganapathisamy "Hybrid Sampling and Similarity Attention Layer in Bidirectional Long Short Term Memory in Credit Card Fraud detection", *International Journal of Intelligent Engineering and Systems*, Vol.15, No.6, 2022.

[21] S. Y. Huang, C.-C. Lin, A.-A. Chiu, and D. C. Yen, "Fraud detection using fraud triangle risk factors", *Inf. Syst. Front.*, Vol. 19, No. 6, pp. 1343-1356, 2017, doi: 10.1007/s10796-016-9647-9.

[22] M. Zur Muehlen and R. Shapiro, "Business Process Analytics", *Handbook on Business Process Management 2*, pp. 243-263, 2015, doi: 10.1007/978-3-642-45103-4_10.

[23] S. Huda, R. SarNo. and T. Ahmad, "Increasing Accuracy of Process-based Fraud Detection Using a Behavior Model", *Int. J. Softw. Eng. Its Appl.*, Vol. 10, No. 5, pp. 175-188, 2016, doi: 10.14257/ijseia.2016.10.5.16.

[24] A. Van Looy and A. Shafagatova, "Business process performance measurement: a structured literature review of indicators, measures and metrics", *SpringerPlus*, Vol. 5, No. 1, p. 1797, 2016, doi: 10.1186/s40064-016-3498-1.

[25] S. Srivastava and D. R. Bhatnagar, "Process Mining Techniques for Detecting Fraud in Banks: A Study", 2021.

[26] C. Moreira, E. Haven, S. Sozzo, and A. Wichert, "Process mining with real world financial loan applications: Improving inference on incomplete event logs", *PLOS ONE*, Vol. 13, No. 12, p. e0207806, 2018, doi: 10.1371/journal.pone.0207806.

[27] T. Wibowo and A. N. L. Tobing, "The Implementation of Fraud Risk Assessment and Anti-Fraud Strategy in Government Institution XYZ", *Asia Pac. Fraud J.*, Vol. 6, No. 2, p. 285, Jan. 2022, doi: 10.21532/apfjournal.v6i2.232.

[28] B. Omair and A. Alturki, "A Systematic Literature Review of Fraud Detection Metrics in Business Processes", *IEEE Access*, Vol. 8, pp. 26893-26903, 2020, doi: 10.1109/ACCESS.2020.2971604.

[29] G. Baader and H. Krcmar, "Reducing false positives in fraud detection: Combining the red

flag approach with process mining", *Int. J. Account. Inf. Syst.*, Vol. 31, pp. 1-16, 2018, doi: 10.1016/j.accinf.2018.03.004.

[30] S. Lagraa and R. State, "Process mining-based approach for investigating malicious login events", in *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, pp. 1-5, 2020, doi: 10.1109/NOMS47738.2020.9110301.

[31] R. SarNo. R. D. DewandoNo. T. Ahmad, M. F. Naufal and F. Sinaga, "Hybrid Association Rule Learning and Process Mining for Fraud Detection", *IAENG International Journal of Computer Science*, Vol. 42, No. 2, 2015.