Fraud Detection on Event Log of Bank Financial Credit Business Process using Hidden Markov Model Algorithm

Dewi Rahmawati, Riyanarto Sarno, Chastine Fatichah, Dwi Sunaryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
dewi16@mhs.if.its.ac.id, riyanarto@if.its.ac.id, chastine@cs.its.ac.id, dwi@if.its.ac.id

***Abstract*—Criminal cases of banks have risen by 55% in 2016. One of the reasons is the fraud in business processes cannot be detected early. Responding to that issue, this research proposes a method for detecting fraud on business processes in the bank credit application. This method uses Hidden Markov Models and activity information recorded in the event log. Hidden Markov Model that used for calculating probability possibility of fraud based on the event log. The results show that HMM method can detect fraud appropriately. The experimental results also show that the accuracy of the results is 94%.**

Keywords—bank; event logs; financial; fraud; hidden markov model; process mining

I. INTRODUCTION

A bank has several business processes. The business processes are used to manage the bank [1]. The bank has many types of business processes. One of the types is bank financial credit business process. Bank financial credit business processes can be changed dynamically. This change is due to increased demand and information, changing of market conditions, or policy changes on the bank's business processes [2]. In bank credit, fraud often occurs that causes a loss to the bank. Fraud in Indonesia is currently handled by an agency called the Financial Services Authority (FSA). Financial Services Authority declared a criminal offense banking since 2014 to present is still dominated by the case of credit and engineering of bookkeeping. The case increasingly rising it indicates the banks should improve the management in order not to undermine the reputation as an institution of trust. Based on the statistical handling of the Crime of Banking (banking criminal acts), types of criminal cases banking in 2014 until the third quarter of 2016 is a case of credit with a percentage of 55 per cent, followed by modified recording 21 percent, embezzlement of funds 15 percent, fund transfers 5 percent and acquisitions 4 percent, As in the world there are 1,388 fraud caused losses of 1.4 billion US dollars in 96 countries.

Fraud can occur due to violations of business processes or standard operating procedures and manipulation of data. Fraud can be defined as a crime that using deception as the main mode and includes a variety of irregularities by individuals and organizations. Fraud, if not prevented and detected will cause a great loss to the bank. Therefore, research is needed on techniques that can be used to detect fraud in business processes [3].

Generally, there are two techniques that can be used to detect fraud in business processes. Such techniques are process mining and data mining techniques. In previous research, the process of mining techniques has contributed to reduce internal fraud on business processes. The research uses several techniques such as conformance checker mining process, analysis with dotted chart, miner with social network, task matrix, and others to conduct an investigation into the event logs generated from business processes running [4, 5].

In this research, the data obtained from the event log on each event log data on bank credit application. The event log data then analyzed using an algorithm Hidden Markov Model (HMM) In order to rebuild process model (a model that produces activity traces) and detect fraud symptoms. The analysis process is particularly useful for detecting anomalies on business processes that occur during the process of the Bank Credit Application.

II. RESEARCH METHOD

A. Bank Financial Credit

Bank Financial Credit is fiscal adroitness permit a someone or commercial enterprise to adopt money and give it back inside the prescribed periods [6]. UU no. 10 of 1998 mentions that the credit is the provision of money or bills can be equated with it, based on agreements between bank lending with another party that obliges the borrowers to refund charge after a certain period of time and gives interest. If someone uses a credit, then it will be subject to interest charges. When banks lend money to customers, the bank certainly expect his money back [2, 5, 7].

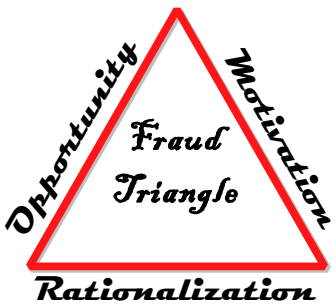

Fig. 1. Fraud Triangle.

To minimize risk (money is not turned back, for example), to provide credit bank should consider a few things related to goodwill (willingness to pay) and the ability to pay (capability to pay) customers to pay off back loan with giving interest. Those things are made up of character (personality), capacity (capacity), capital (capital), collateral (warranties), and condition economy (economic conditions) as 5C (penta C) [8, 9, 10].

B. Fraud

Based on the definition of the Institute of Internal Auditors, fraud is "An arrangement of capriciousness and savage acts describe intentional deception" which means "set of actions that are not permitted and unlawful characterized by an element of fraud intentional" [11, 12]. The Association of Certified Fraud Examiners (ACFE) or the Association of Audit Fraud Certified) divides fraud into three types namely deviations on assets (asset Misappropriation), Statement of Counterfeit (Fraudulent Statement), and corruption (Corruption) [13].

There are three things that contribute to a fraud attempt, the pressure (boost), opportunities (opportunities), and rationalization (rationalization) as in [14] illustrated in Fig. 1.

C. Event Logs

Event log is a recording process in the form of transaction history or audit trail in an information system tool as in [15]. Each system event log information definitely has evidence of ongoing transactions. For example only existing recording of the event log on the Enterprise Resource Planning (ERP). Event Log contains information about activities in the form of a case or a specific task [16].

The case itself is called the "process instance" is an ongoing activity. For example, the order to the supplier (purchasing), the order by the customer (customer order) and several other events [17, 18]. While the task is activated in the trace, could be stages of activity. So in the trace can have many tasks. Event Log consists of several attributes including the Case ID, task, event, user (originator) and time (timestamp).

D. Hidden Markov Model

Associations of the two processes in which one process has a hidden variable called the state and the determination of these variables by other processes that produce random variable circuit or circuit observer [19, 20].

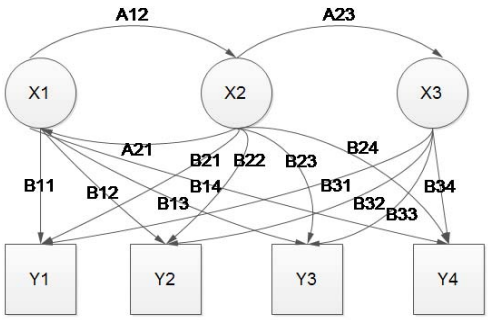

Fig. 2. Probabilistic parametric quantity Hidden Markov Model.

In a Hidden Markov model, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly.An example of Hidden Markov Models can be seen in Fig. 2.

III. PROPOSED METHOD

A. Data Source

The data obtained by the authors is the data of the financial system in the form of bank data division activities performed and data on bank employees in the form of data in .csv format. The data in .csv format consists of several columns containing:

1. Event_id: contains activities that have been carried out
2. Case_id: contains user id procurement had been forged
3. Activity: contains the name of the activity that has been performed
4. Start_Time: contains the date and start time activity
5. Complete_Time: contains the date and time of completion activities
6. Resource: contains the user who perform these activities

B. Procedure Research

The procedures in this research starting from data collection event log .csv format. Then use the .csv formatted log event data to calculate the value probability matrix algorithms using hidden Markov models. After getting probability value matrix, checking the data which it contains elements of fraud or not. Fraud will be detected if the state probability of fraud is greater than the value of state probability of no fraud, and vice versa, if the value of state probability no fraud is higher than the probability of fraud, the state is not detected symptoms fraud case.

C. Design Workflow

Workflow design of bank credit application was originally created using .pnml of Woped Editor format as illustrated in Fig. 3. Woped Editor is a tool that is used to create workflow logic that describes.

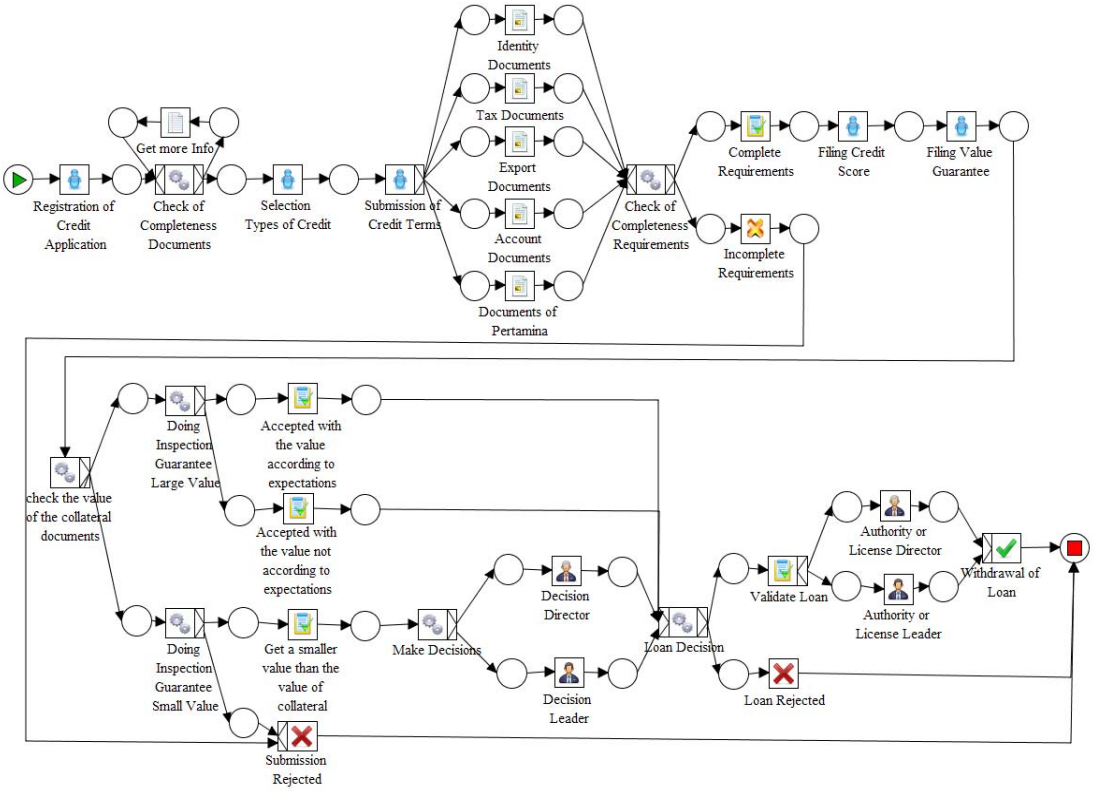

Fig. 3. Workflow design of bank credit application.

TABLE I. TABLE EVENT LOG BANK FINANCIAL CREDIT BUSINESS PROCESS

Event ID	Case ID	Activity	Start Time	Complete Time	Resource
1	1	Register with Credit Application	12/27/16 13:00	12/27/16 13:45	Clerk
2	1	Check of Completeness Documents	12/27/16 13:45	12/27/16 14:30	Clerk
3	1	Get More Info	12/27/16 14:30	12/27/16 15:15	Clerk
4	1	Selection of Types of Credit	12/27/16 15:15	12/27/16 16:00	Clerk
5	1	Submission of Credit Terms	12/27/16 16:00	12/27/16 16:45	Clerk
6	1	Identity Documents	12/27/16 16:45	12/27/16 17:30	Clerk
7	1	Tax Documents	12/27/16 17:30	12/27/16 18:15	Clerk
8	1	Export documents	12/27/16 18:15	12/27/16 19:00	Clerk
9	1	Account Documents	12/27/16 19:00	12/27/16 19:45	Clerk
10	1	Documents of Pertamina	12/27/16 19:45	12/27/16 20:30	Clerk
11	1	Check of Completeness Requirements	12/27/16 20:30	12/27/16 21:15	Clerk
12	1	Complete Requirements	12/27/16 21:15	12/27/16 22:00	Clerk
13	1	Filing Credit Score	12/27/16 22:00	12/27/16 22:45	Clerk
14	1	Filing Value Guarantee	12/27/16 22:45	12/27/16 23:30	Clerk
15	1	Checks Security Documents / Collateral Value	12/27/16 23:30	12/28/16 0:15	Manager
16	1	Doing Inspection Guarantee Large Value	12/28/16 0:15	12/28/16 1:00	Staff
17	1	Accepted Rated accordance expectation	12/28/16 1:00	12/28/16 1:45	Staff
18	1	Decision Loans	12/28/16 1:45	12/28/16 2:30	Manager
19	1	Doing Validation Loans	12/28/16 2:30	12/28/16 3:15	Manager
20	1	Authority or License Director	12/28/16 3:15	12/28/16 4:00	Manager
21	1	Authority or Permit Leaders	12/28/16 4:00	12/28/16 4:45	Manager
22	1	Withdrawal of Loan	12/28/16 4:45	12/28/16 5:30	Manager

TABLE II. DETERMINING NO FRAUD VALUE VIEWS FROM 6 CRITERIA

No Fraud	Skip Sequence	Bad	1	0.01
	Skip Decision	Very Weak	2	0.03
	Wrong Throughput Time (Min)	Very Weak	2	0.03
	Wrong Throughput Time (Max)	Very Weak	2	0.03
	Wrong Resource	Very Weak	2	0.03
	Wrong Duty Sequence	Very Weak	2	0.03
	Wrong Duty Decision	Very Weak	2	0.03
	Wrong Duty Combine	Very Weak	2	0.03
	Wrong Pattern	Very Weak	2	0.03
	Wrong Decision	Very Weak	2	0.03
	WFC Small	Weak	3	0.04
	WFC Medium	Middle	11	0.14
	WFC Large	Strong	12	0.16
	No Skip	Very Strong	15	0.19
	Under Collateral	Very Weak	2	0.03
	Over Collateral	Very Strong	15	0.19
	Skip Sequence	Bad	1	0.01
	Skip Decision	Very Weak	2	0.03

working with existing business process flows. The Workflow represents the flow of bank credit. Then to get the Event Log data sets which later will analyzed by the Hidden Markov Model, a process model are simulated like in Table I.

IV. RESULT AND ANALYSIS

The experiments for this research is a bank loan application process. This research using event log with 90 cases to determine the activities are influenced by fraud.

This paper uses 16 indicators or ways to detect fraud. The indicator i.e. "Skip Activity Decision", "Skip Activity Sequence", "Wrong Resource", "Wrong Throughput of Time (Max)", "Wrong Throughput of Time (Min)", “Wrong Decision", "Wrong Duty Decision", "Wrong Pattern", "WFC Small", "WFC Medium", "WFC Large", "No Skip", "Wrong Duty Sequence", "Under Collateral", "Wrong Duty Combine", and "Over Collateral".

The first indicator is "skip sequence" which means the process straight passes. The second indicator "skip decision" which means the decision process are passes, the third "wrong throughput time (min)" which means the process with less time than the standard prescribed time, the fourth "wrong throughput time (max)" which means the process with a longer time than the standard prescribed time, the fifth "wrong resource" which means resource is not supposed to do the work, the sixth "wrong duty sequences” which means when in one case there is an actor are executes two or more activity types of sequences simultaneously, the seventh "wrong duty decision" which means when in one case there is an actor are executes two or more activity type of decision at the same time, Eighth is "duty combine wrong" which means when in one case there is an actor are executes two or more types of activity decision and sequences simultaneously. The tenth wrong decision which means

TABLE III. DETERMINING FRAUD VALUE VIEWS FROM 6 CRITERIA

Fraud	Skip Sequence	Very Strong	15	0.08
	Skip Decision	Strong	14	0.08
	Wrong Throughput Time (Min)	Middle	13	0.07
	Wrong Throughput Time (Max)	Strong	14	0.08
	Wrong Resource	Strong	14	0.08
	Wrong Duty Sequence	Strong	14	0.08
	Wrong Duty Decision	Strong	14	0.08
	Wrong Duty Combine	Strong	14	0.08
	Wrong Pattern	Strong	14	0.08
	Wrong Decision	Strong	14	0.08
	WFC Small	Strong	14	0.08
	WFC Medium	Weak	4	0.02
	WFC Large	Very Weak	3	0.02
	No Skip	Bad	1	0.01
	Under Collateral	Strong	14	0.08
	Over Collateral	Bad	1	0.01
	Skip Sequence	Very Strong	15	0.08
	Skip Decision	Strong	14	0.08

TABLE IV. STATE TRANSITION MATRIX FRAUD

State Transition Matrix		
	Fraud	***No Fraud***
Fraud	0.7	0.3
No Fraud	0.3	0.7

the condition when an activity type of decision contains errors of decision-making. Eleventh "WFC Small" which means the Workflow Fragment Component contains one document of activity submission credit terms (load activities XOR), the twelve "WFC Medium" which means the Workflow Fragment Component contains two documents of activity submission credit terms (load activities AND), third twelve "WFC Large" which means the Fragment Workflow Component contains 3-5 document handing over credit terms (load activities OR), Fourteenth "No skip" which means a correct sequence of activities and are not exceeded, the fifteenth "Under Collateral" which means the value of loan collateral is smaller than the money borrowed, the sixteenth "Over Collateral" which means the value of loan collateral is smaller than the money borrowed.

First, this paper determines the weights of indicators of fraud and no fraud using the algorithm of HMM. The weights of the indicators are shown in Table II, and Table III. Secondly, this paper determines state transition matrix of fraud and no fraud as in Table IV. Thirdly, this paper calculates the value of the observation probability matrix of values that have been defined in Table II, and II as in Table V. The fourth process defines the initial state probability of fraud as in Table VI.

TABLE V. OBSERVATION MATRIX FRAUD

	Type of Anomaly		Fraud	No Fraud
Observation Probability Matrix	***Skip***	***Sequence***	0.08	0.01
		Decision	0.08	0.03
	Wrong Throughput Time	***Min***	0.07	0.03
		Max	0.08	0.03
	Wrong Resource		0.08	0.03
	Wrong Duty	***Sequence***	0.08	0.03
		Decision	0.08	0.03
		Combine	0.08	0.03
	Wrong Pattern		0.08	0.03
	Wrong Decision		0.08	0.03
	WFC Small		0.08	0.04
	WFC Medium		0.02	0.14
	WFC Large		0.02	0.16
	No Skip		0.01	0.19
	Under Collateral		0.08	0.03
	Over Collateral		0.01	0.19

TABLE VI. PROBABILITY INITIAL STATE FRAUD

Probability Initial State	***Fraud***	0.5
	No Fraud	0.5

TABLE VII. FRAUD DETECTION RESULTS IN 1 CASE

Type of Anomaly	Before Normalize		Normalize		State Chosen
	Fraud	*No Fraud*	*Fraud*	*No Fraud*	
No Skip	0.002	0.09	0.02	**0.97**	**No Fraud**
Skip Sequence	0.02	0.008	**0.73**	0.26	**Fraud**
Skip Sequence	0.006	0.002	**0.73**	0.26	**Fraud**
No Skip	0.0004	0.03	0.01	**0.98**	**No Fraud**
WFC Large	0.005	0.10	0.04	**0.95**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**

One of the processes in an event log which are used in this experiment is Register with Credit Application (No Skip) -> Check Completeness Document (Skip Sequence) -> More Information (Skip Sequence) -> Selection Types of Credit (No Skip) -> Submission of Credit Terms (WFC Large) -> Document Identity (No Skip) -> Tax Document (No Skip) -> Export Document (No Skip) -> Documents Account (No Skip) -> Document Pertamina (No Skip) -> Check completeness Requirements (No Skip) -> Requirements Complete (No Skip) -> Filing Credit Score (No Skip) -> Filing Value Guarantee (No Skip) -> Check Value Security Documents / Collateral

TABLE VIII. FRAUD DETECTION RESULTS IN 1 CASE (CONT.)

Type of Anomaly	Before Normalize		Normalize		State Chosen
	Fraud	*No Fraud*	*Fraud*	*No Fraud*	
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
Over Collateral	0.001	0.13	0.01	**0.98**	**No Fraud**
Over Collateral	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**
No Skip	0.001	0.13	0.01	**0.98**	**No Fraud**

TABLE IX. ACCUMULATION FRAUD VALUE FROM 90 CASE

Type of Anomaly	Number of Case	Number of Activities	State Probability Value Fraud	State Probability Value No Fraud
Skip Sequence	1	4	0.73	0.26
Skip Decision	1	6	0.56	0.43
Wrong Throughput Time Min	1	1	0.54	0.45
Wrong Throughput Time Max	1	1	0.56	0.43
Wrong Resource	1	1	0.56	0.43
Wrong Duty Sequence	1	1	0.56	0.43
Wrong Duty Decision	1	3	0.56	0.43
Wrong Duty Combine	1	4	0.56	0.43
Wrong Pattern	1	8	0.56	0.43
Wrong Decision	1	2	0.56	0.43
WFC Small	1	1	0.46	0.23
WFC Medium	1	1	0.06	0.93
WFC Large	78	78	0.04	0.95
No Skip	90	1695	0.01	0.98
Under Collateral	1	2	0.56	0.43
Over Collateral	80	158	0.01	0.98

(Over Collateral) -> Conduct Examination Value Guarantee Great (Over Collateral) -> Received Rated accordance Hope (No Skip) -> Decision Loans (No Skip) -> Perform Validation Loans (No Skip) -> Authority or License Director (No Skip) -> Authority or Permit Leader (No Skip) -> Withdrawal Loan (No Skip). The result is show in Table VII, and Table VIII.

TABLE X. TABLE ACTUAL AND OBTAIN ACTIVITY

		Actual Activity	
		True	***False***
Obtained Activity	***True***	80	0
	False	10	0

TABLE XI. THE QUALITY OF HIDDEN MARKOV MODEL

Fraud Detection of Bank Financial Credit Business Process using	Recall	Precision	F-Score
Hidden Markov Model	1	0.89	0.94

This research takes 90 cases wherein 10 cases are detected fraud and 80 cases are detected no fraud as written in Table X. All of the cases will be processed by the method using Hidden Markov Model. Fraud is detected based on 16 categories of activities. The categories are written in Table IX and the accuracy Fraud Detection of Bank Financial Credit Business Process using Hidden Markov Model is written in Table XI.

Fraud will be detected if the state probability of fraud is greater than the value of state probability of no fraud, and vice versa, if the value of state probability no fraud is higher than the probability of fraud, the state is not detected symptoms fraud case.

V. CONCLUSION

This experiment identifies 90 cases of bank loan applications event logs. The cases contain 10 cases of fraud and 80 cases of no fraud. This paper concludes that the accuracy of the experiment by using Hidden Markov Modell is 94%.

The experiment results also showed that algorithm utilized Hidden Markov Model can obtain cases which detected symptoms of fraud in the testing data.

REFERENCES

[1] K.-D. Gronwald, “ERP: Enterprise Resource Planning,” Integrated Business Information Systems, pp. 59–86, 2017.

[2] S.N. Chary, Production and Operations Management, New Delhi: Tata McGraw-HillEducation, 2009.

[3] O. Monisola, "Effect of Internal Audit On Prevention of Frauds, Errors And Irregularities In Corporate Organisation," in IISTE, Nigeria, 2013.

[4] W.V.D. Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Netherlands: Springer, 2011.

[5] A. Rozinat, and W. DerAalst, Conformance Checking of Processes Based on Monitoring Real Behavior, Netherlands: Eindhoven University of Technology, 2007.

[6] I. Merriam-Webster, "Merriam Webster," Merriam Webster Incorporated, 5 March 2015. [Online]. Available: http://www.merriam-webster.com/dictionary/credit. [Accessed 25 December 2016].

[7] A. Kassahun, and B. Tekinerdogan, “Architecture Viewpoint for Modeling Business Collaboration Concerns using Workflow Patterns,” Proceedings of the 11th International Joint Conference on Software Technologies, 2016.

[8] S. Huda, R. Sarno, and T. Ahmad, " Increasing Accuracy of Process-based Fraud Detection Using a Behavior Model," International Journal of Software Engineering and Its Applications (IJSE), pp. 175-188, 2016.

[9] R. Sarno, H. Ginardi, E.W. Pamungkas, and D. Sunaryono, "Clustering of ERP business process fragments," International Conference on Computer, Control, Informatics and Its Applications (IC3INA), p. 319 – 324, 2013.

[10] R. Sarno, Kartini, W.A. Wibowo, and A.S.A, "Time Based Discovery of Parallel Business Processes," in The 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2015.

[11] S. Huda, R. Sarno, T. Ahmad, and H.A. Santoso, "Identification of Process-based Fraud Patterns in Credit Application," in Proceedings International conference on information and communication technology (ICoICT), 2014.

[12] S. Huda, R. Sarno, and T. Ahmad, "Fuzzy MADM Approach for Rating of Process-Based Fraud," Journal of ICT Research and Applications, pp. 111-128, 2015.

[13] R. Sarno, Fernandes, D. Sunaryono, and A. Munif, "Business Process Anomaly Detection using Ontology-based Process Modeling and Multi-level Class Association Rule Learning," in The 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2015.

[14] D. Rahmawati, M.A. Yaqin, and R. Sarno, “Fraud detection on event logs of goods and services procurement business process using Heuristics Miner algorithm,” 2016 International Conference on Information & Communication Technology and Systems (ICTS), 2016.

[15] A. Solti, L. Vana, and J. Mendling, “Time Series Petri Net Models,” Data-Driven Process Discovery and Analysis, pp. 124–141, 2017.

[16] F.-S. Hsieh, “A hybrid and scalable multi-agent approach for patient scheduling based on Petri net models,” Applied Intelligence, May 2017.

[17] D. Patel, and M. Pandya, “Optimizing Resource Conflicts in Workflow Management Systems Using an Innovative Algorithm,” International Journal of Science and Research (IJSR), vol. 5, no. 2, pp. 2100–2103, Feb. 2016.

[18] R. Sarno, F. Haryadita, Kartini, Sarwosri, and A.S.A, "Business Process Optimization from Single Timestamp Event Log," in The 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 2015.

[19] R. Sarno, and K.R. Sungkono, "Hidden Markov Model for Process Mining of Parallel Business Processes," International Review on Computers and Software (IRECOS), pp. 290-300, 2016.

[20] R. Sarno, and K.R. Sungkono, "Coupled Hidden Markov Model for Process Mining of Invisible Prime Tasks," International Review on Computers and Software (IRECOS), pp. 539-547, 2016.