CHMM for Discovering Intentional Process Model From Event Logs By Considering Sequence of Activities

Kelly R. Sungkono
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
kelsungkono@gmail.com

Riyanarto Sarno
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Abstract— **An intentional process model is known to analyze processes deeply and provide recommendations for the upcoming processes. Nevertheless, the discovery of intentions is a difficult task because the intentions are not recorded in the event log, but they encourage the executable activities in the event log. Map Miner is the latest algorithm to depict the intentional process model. A disadvantage of this algorithm is the inability to determine strategies that contain same activities with the different sequence with other strategies. This disadvantage leads failure on the intentional process model. This research proposes an algorithm for discovering an intentional process model by considering the sequence of activities and CHMM (Coupled Hidden Markov Model). The probabilities and states of CHMM are utilized for the formation of the intentional process model. The experiment shows that the proposed algorithm with considering the sequence of activities gets an appropriate intentional process model. It also demonstrates that an obtained intentional process model using proposed algorithm gets the better validity than an intentional process model using Map Miner Method.**

Keywords—Coupled Hidden Markov Model; Event Log; Intention Mining; Process Model; Validity.

I. INTRODUCTION

Nowadays, process modeling has become a concern. A process model can be used for analyzing executable processes in the information systems and understanding the behavior of executor in implementing the processes [1]–[3]. All research about process discovery, e.g. [4]–[7], had proposed algorithms for modeling processes based on the sequence of activities from the event logs. Meanwhile, for understanding the behavior deeply, intentions as the basis of the chosen activities in a process are more needed than only activities [8]. Therefore, it also requires an intentional process model based on the event log.

Intention mining is a study that mines activities for knowing and predicting the intentions that are related to the activities [9]. On analysis side, the obtained intentions can show slits between defined activities in the model and executed activities in the event log [9]. Moreover, in system side, the obtained intentions can provide recommendation acts for the executor of processes. There are several research about intention mining, such as the discovery of consumption intentions of social media [10], the discovery of consumption intentions of information on personal websites [11], user intentions modeling during the creation processes of Entity-Relationship diagrams [9], and developer intentions modeling during the use of Eclipse platform [8]. The latest research about forming an intentional process model is [8]. The name of the algorithm in [8] is Map Miner Method.

Map Miner Method is an algorithm for discovering an intentional process model by utilizing Hidden Markov Model, K-means and several rules. Firstly, this method built Hidden Markov Model with strategies as the states and activities from the event log as the observations. Strategies are steps of the displacement of one intention to another. They contain one or several activities that are recorded in the event log. Map Miner Method chose strategies with high dependency probabilities toward the activities. Then, it utilized K-means with additional rules to process obtained strategies into an intentional process model.

Not all strategies contain different activities with others. Some of them consist of the same activities (but they have different order executions) with other strategies. In this condition, Map Miner Method is difficult to determine the right strategy because this method determines the strategy only based on the probability of each activity. In fact, the fallacy in the selection of strategies has an impact on the wrong intentional process model.

This research proposes an algorithm for discovering an intentional process model by utilizing several rules and CHMM. The rules are used to determine the strategies by considering the sequence of activities. CHMM that has the obtained strategies and activities from the event log as its first and second observation is used to form the intentional

978-1-5386-0549-3/17/$31.00 ©2017 IEEE

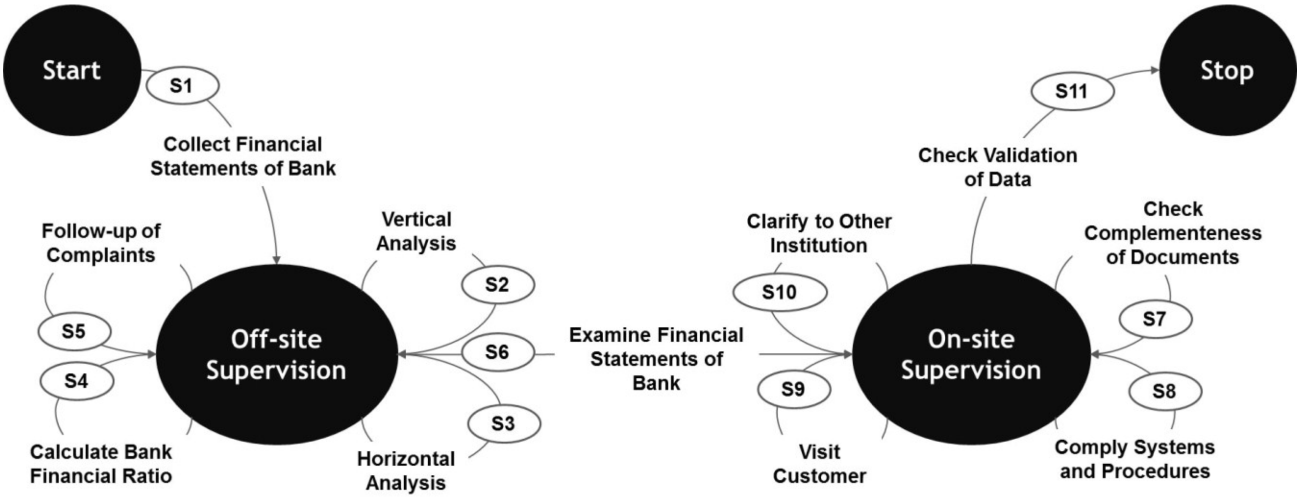

Fig. 1. An intentional process model of Patterns of Fraud Detections

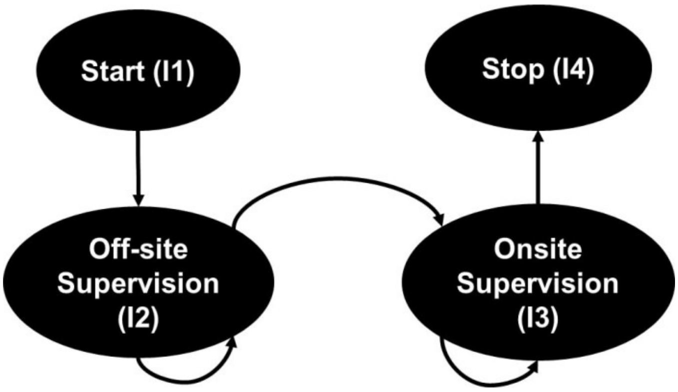

Fig. 2. An general intentional process model of Patterns of Fraud Detections

process model. CHMM has been applied in several issues, i.e. process discovery [4], and bearing fault recognition [12]. This research initiates the use of CHMM in the domain of intention mining.

Each of proposed algorithm must be evaluated to demonstrate the quality of the method. The validity to evaluate process model in [13] was chosen to determine the quality of the proposed algorithm of this research. The quality of proposed algorithm was compared with Map Miner Method [8].

II. LITERATURE REVIEW

A. Intentional Process Model

[8] introduces a map model to illustrate an intentional process model. A map model is a graphical model wherein the nodes denote intention and the arcs denote strategies or relations of intentions. Fig.1 shows the example of an intentional process model and the general model of Fig.1 is shown in Fig.2. Fig.2 is utilized as an actual intentional process model in this experiment.

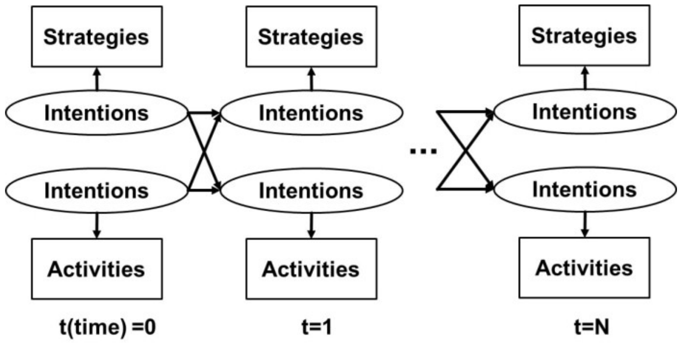

Fig. 3. Proposed CHMM.

Fig.1 describes patterns of fraud detections by the Financial Services Authority [14]. There are two intentions: off-site supervision and on-site supervision. The off-site supervision is a fraud detection by utilizing financial details. The on-site supervision is a fraud detection by observing the site work. There are eleven strategies that can be selected to fulfill the fraud detections. Start and Stop are additional intentions to signify the beginning and the end of patterns of fraud detections.

An explanation of the strategy in Fig.1 is as follows. Strategy *Collect Financial Statements of Banks* fulfills off-site supervision if the current position is Start. Four strategies are determined to fulfill off-site supervision if the current position is also off-site supervision and strategy *Examine Financial Statements of Bank* is determined to change intention from off-site supervision to intention from on-site supervision. Finally, four remaining strategies fulfill on-site supervision if the current position is same supervision and one strategy to achieve the stop intention as the end point of the processes.

The Financial Services Authority is not directly applying the strategies, but execute the activities that perform the strategies. Those activities will be recorded in the event logs. There are 25 activities to perform the strategies of fraud detections. The details of activities and the relations to the strategies are explained in section Result and Analysis.

B. CHMM

CHMM is a compound of two Hidden Markov Model [4]. The states of CHMM rely on the states at previous time. The mathematic form CHMM is outlined in Equation (1). Probabilities of beginning states are denoted by π. A denotes a transition matrix and B denotes two emission matrices. The detailed explanation can be studied in [4].

$$CHMM = (\pi, A, B) \tag{1}$$

The proposed CHMM that is shown in Fig.3 is used for designing an intentional process model. There are three parts in this model: the intentions as the states, the strategies as the observations, and the activities as the second observations. This CHMM is trained by Baum-Welch method in [12] and [4].

C. The Validity of Model

Each method can be evaluated based on its output. The proposed algorithm was evaluated by comparing the quality of its obtained model with the quality of obtained model by Map Miner Method. The validity is chosen in this paper to determine the quality of those obtained models and it has been applied for evaluating a process model [13].

A model is categorized as valid if all statements in the model are correct. Statements are the data that can be obtained based on the model. In the intentional process model, the statements are intentions and dependencies between intentions. The statements are correct if those are same with the statements of an actual intentional process model.

To define the validity of the models, this paper compares Causal Nets [13] that construct intentions and those linkages with others based on the models of proposed algorithm or Map Miner Method [8] with Causal Nets based on the actual model which is shown in Fig. 2.

III. PROPOSED ALGORITHM

A method has been proposed to discover a general intentional process model. This method is a combination of three general steps. Firstly, strategies are obtained by considering sequences of activities. Afterwards, CHMM is constructed using the obtained strategies and activities in event logs. Lastly, An intentional process model is depicted based on probability matrices of CHMM.

The proposed algorithm is described in TABLE I. Steps from step 1 to step 19 tell about the steps for obtaining strategies which are stored in list chosen_strag. The strategies are chosen by their probabilities which are stored in list S_prob. The probability of strategy (*S_prob[s]*) is higher if the activity in the strategy is the first activity (*position = 0*) or the sequence of activities in the event log is same as the sequence of activities that perform the strategy. The strategy which has highest probability is stored in list chosen_strag when the next activity in the event log is indicated to perform another strategy or *check_strag* is zero. Baum Welch method in step 20 is derived from [4]. This aim of Baum Welch method is training the CHMM. Before Baum-Welch method was applied, the matrices of CHMM were initialized. All values in an initial transition matrix (*A_first*) were 1/numberofcolumns and the probability of beginning state (*π_first*) is 1. Afterwards, emission matrices (*B_first*) took values based on the execution of activities in processes and obtained strategies. For the first emission matrix, first strategies of start intention and last strategies of end intention were initialized to 1. Each strategy which had relation to first strategies depended on next intention of the start intention and each strategy which had relation to last strategies depended on previous intention of the stop intention. Elements of the emission matrix that illustrated those dependencies were numbered at 1. Subsequently, other elements that had not been given values were numbered at 0,1. The second emission matrix implemented initialization steps of first emission matrix with activities as its input. Finally, the model was depicted by constructing the relation of intentions that has more than zero probability (*a(int_bef,int_aft) > 0*) in the transition matrix.

TABLE I. PROPOSED ALGORITHM

A Method of Discovering A General Intentional Process Model	
Input : Actv (list of activities in the event log), S (list of strategies)	
1	chosen_stag ← (empty list)
2	position = 0
3	for a in Actv then
4	S_prob ← (empty list)
5	check_strag = 0
5	for s in S then
6	if a in s[position] and (position = 0 or the previous activity of a in s[position-1]) then
7	S_prob[s].add(1)
8	check_strag ← check_strag + 1
8	endif
9	endfor
10	position ← position + 1
11	if check_strag = 0 then
12	chosen_strag.insert(s which has maximum value in S_prob)
13	position = 0
14	*repeat step 4 until step 10*
15	endif
16	if a is last_activity then
17	chosen_strag.insert(the last s in list S)
18	endif
19	endfor
20	*π, A, B* = BaumWelch (*π_first, A_first, B_first*, Actv, chosen_strag)
21	for *a*(int_bef,int_aft) in A then
22	if *a*(int_bef,int_aft) > 0 then
23	*add relation from intention int_bef to intention int_aft in a model*
24	endif
25	endfor

TABLE II. LIST OF ASSOCIATED ACTIVITIES WITH STRATEGIES

Strategies		Associated Trace Activities	The Codes of Activities
S1	Collect Financial Statements of Bank	Define Bank Financial Report	AFD1
S2	Vertical Analysis	Check Trend Transaction, Check Credit Surge, Check Lowering of Interest, **Check Completeness Bill of Sales, Check Completeness Marketable Securities**	AFD2, AFD3, AFD4, **AFD 19, AFD20**
S3	Horizontal Analysis	Check Financial Disbursement, Check Decrease Financial Value	AFD5, AFD6
S4	Calculate Bank Financial Ratio	Calculate Liquidity, Calculate Modal Structure and Solvability, Calculate Return on Investment, Calculate Profile Margin, Calculate Assets Utilization, Calculate Market Measure	AFD7, AFD8, AFD9, AFD10, AFD11, AFD12
S5	Follow-up of Complaints	Define Follow-up complaints, Specify Actions Related with the Complaints	AFD13, AFD14
S6	Examine Financial Statements of Bank	Check Company Profit, Check Sales, Check Dividend, Check Equity	AFD15, AFD16, AFD17, AFD18
S7	Check Completeness of Documents	Check Completeness Bill of Sales, Check Completeness Marketable Securities, **Check Trend Transaction, Check Credit Surge, Check Lowering of Interest**	AFD19, AFD20, **AFD2, AFD3, AFD4**
S8	Comply Systems and Procedures	Check Conformance Procedures	AFD21
S9	Visit Customer	Define Follow-up a Visit to Customer, Define Bank Financial Reports	AFD22, AFD1
S10	Clarify to Other Institution	Check Financial History in Other Bank, Check Transaction History with Other Relations	AFD23, AFD24
S11	Check Validation of Data	Financial Statements Valid	AFD25

IV. RESULT AND ANALYSIS

A. Data

Fig. 2 is an actual intentional process model that utilized in the experiment, wherein the nodes (circles) are the intentions and the arcs are relations of those intentions. The goal of this paper is to get an intentional model of fraud detections that resembled Fig. 2.

TABLE III. A PIECE OF EVENT LOG

Case ID	Activities	Start Stamp	Case ID	Activities	Start Stamp
PP1	AFD1	2/16/2016 10:32	PP1	AFD11	2/16/2016 16:52
PP1	AFD2	2/16/2016 13:42	PP1	AFD12	2/16/2016 16:52
PP1	AFD3	2/16/2016 15:17	PP1	AFD15	2/16/2016 16:52
PP1	AFD4	2/16/2016 16:04	PP1	AFD16	2/16/2016 16:52
PP1	AFD19	2/16/2016 16:28	PP1	AFD17	2/16/2016 16:52
PP1	AFD20	2/16/2016 16:40	PP1	AFD18	2/16/2016 16:52
PP1	AFD5	2/16/2016 16:46	PP1	AFD19	2/16/2016 16:52
PP1	AFD6	2/16/2016 16:49	PP1	AFD20	2/16/2016 16:52
PP1	AFD7	2/16/2016 16:50	PP1	AFD2	2/16/2016 16:52
PP1	AFD8	2/16/2016 16:51	PP1	AFD3	2/16/2016 16:52
PP1	AFD9	2/16/2016 16:51	PP1	AFD4	2/16/2016 16:52
PP1	AFD10	2/16/2016 16:52	PP1	AFD25	2/16/2016 16:52

TABLE IV. THE CHOSEN CORRESPONDING STRATEGIES

Identification of Case	The activities	The corres-ponding strategies	The activities	The corres-ponding strategies
PP1	AFD1	S1	AFD15, AFD16, AFD17, AFD18	S6
	AFD2, AFD3, AFD4, AFD19, AFD20	S2	AFD19, AFD20, AFD20, AFD2, AFD3, AFD4	S7
	AFD5, AFD6	S3	AFD25	S11
	AFD7, AFD8, AFD9, AFD10, AFD11, AFD12	S4		

To prove that the strength of considering the sequence of activities, several strategies with their related activities from The Financial Service Authority were modified. The modified strategies were strategy S2 and strategy S7. The modification was performed by adding activities of a strategy to other strategy. Bold text of TABLE II shows the result of the modification.

This paper used an event log with 50 processes. The processes were simulation data that were formed according to the actual intentional model of fraud detections and a modified list of strategies and related activities in TABLE II. Case PP1 which is shown in TABLE III is the example of processes in the event log.

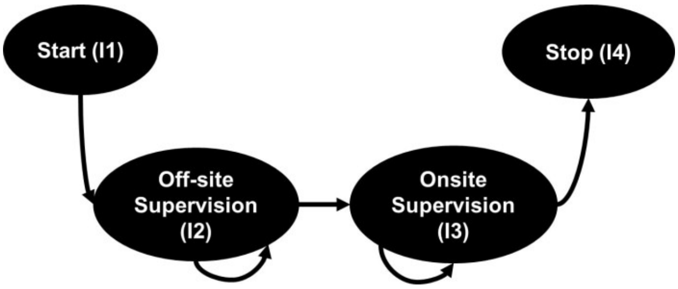

Fig. 4. Discovered General Intentional Process Model of Patterns of Fraud Detections using Proposed algorithm

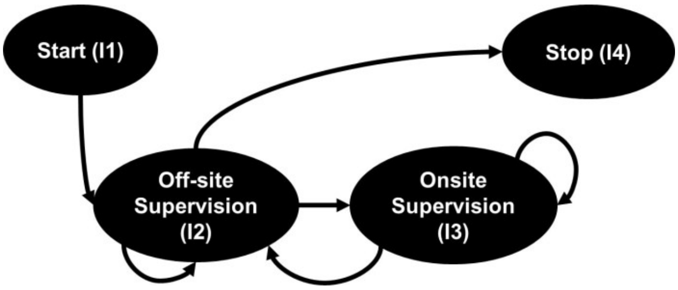

Fig. 5. Discovered General Intentional Process Model of Patterns of Fraud Detections using Map Miner Method

B. Experiment

Using data in Section IV.A and Fig.2, this experiment is carried out by comparing the result of proposed algorithm and the result of Map Miner Method. The steps of proposed algorithm followed steps in Section III and the steps of Map Miner Method followed steps in [8]. The deduction of chosen strategies as the result of the first step in proposed algorithm is described in TABLE IV.

The intentional process models of proposed algorithm and Map Miner Method are illustrated in Fig. 4 and Fig. 5. To measure the validity of the discovered intentional process models, those models and the actual model were converted into Causal Nets. TABLE V until TABLE VII shows the Nets.

As shown in TABLE V, TABLE VI, and TABLE VII, all elements of Causal Nets of an intentional model by proposed algorithm are same as those of the actual intentional model. Otherwise, only the leading intentions of Start and On-site and the following intentions of Start and Stop in Causal Nets by Map Miner Method are same as Causal Nets of the actual intentional model. It is proven that only an intentional process model of proposed method is a valid intentional model.

TABLE V. THE CAUSAL NET OF ACTUAL INTENTIONAL MODEL

Leading intentions	Intentions	Following intentions
{Ø}	Start	{{Off-site}}
{{Start,Off-site}}	Off-site	{{Off-site,On-site}}
{{Off-site,On-site}}	On-site	{{On-site,Stop}}
{{On-site}}	Stop	{{ Ø }}

TABLE VI. THE CAUSAL NET OF INTENTIONAL MODEL USING PROPOSED ALGORITHM

Leading intentions	Intentions	Following intentions
{Ø}	Start	{{Off-site}}
{{Start,Off-site}}	Off-site	{{Off-site,On-site}}
{{Off-site,On-site}}	On-site	{{On-site,Stop}}
{{On-site}}	Stop	{{ Ø }}

TABLE VII. THE CAUSAL NET OF INTENTIONAL MODEL USING MAP MINER METHOD

Leading intentions	Intentions	Following intentions
{Ø}	Start	{{Off-site}}
{{Start, Off-site, On-site}}	Off-site	{{Off-site,On-site, Stop}}
{{Off-site,On-site}}	On-site	{{On-site,Off-site}}
{{Off-site}}	Stop	{{ Ø }}

V. CONCLUSION

This paper discovers the intentional process model automatically based on activities of the event log. This paper gives a proposed algorithm utilizing Coupled Hidden Markov Model (CHMM) and considering sequence of activities.

The proposed algorithm is divided into three steps. Firstly, a method of obtaining strategies by considering sequence of activities was applied. Afterwards, constructed CHMM using the obtained strategies and activities in event logs. Lastly, depicted an intentional process model based on probability matrices of CHMM.

The outcomes of the experiment show that the proposed algorithm can illustrate the right intentional model of patterns of fraud detections. Furthermore, those also show that an obtained intentional process model by proposed algorithm has better validity than the obtained intentional process model by Map Miner Method.

This proposed algorithm is promising because this is a paper that utilizes CHMM in the domain of intention mining. The future work of this research is applying the proposed algorithm in the large-scale real event logs.

ACKNOWLEDGMENT

Authors give a deep thank to Institut Teknologi Sepuluh Nopember and the Ministry of Research, Technology and Higher Education of Indonesia for supporting the research.

REFERENCES

[1] S. Huda, R. Sarno, T. Ahmad, and H. A. Santoso, "Identification of Process-based Fraud Patterns in Credit Application," in *Information and Communication Technology (ICoICT), 2014 2nd International Conference on*, 2014, pp. 84–89.

[2] 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.

[3] R. Sarno, E. W. Pamungkas, D. Sunaryono, and Sarwosri, "Business Process Composition Based on Meta Models," in *Intelligent Technology and Its Applications (ISITIA)*, 2015, pp. 315–318.

[4] R. Sarno and K. R. Sungkono, "Coupled Hidden Markov Model for Process Mining of Invisible Prime Tasks," *IRECOS (International*

Rev. Comput. Software), vol. 11, no. 6, pp. 539–547, 2016.

[5] R. Sarno and K. R. Sungkono, “Hidden Markov Model for Process Mining of Parallel Business Processes,” *IRECOS (International Rev. Comput. Software)*, vol. 11, no. 4, pp. 290–300, 2016.

[6] R. Sarno, W. A. Wibowo, Kartini, and F. Haryadita, “Determining Model Using Non-Linear Heuristics Miner and Control-Flow Pattern,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 1, pp. 349–360, 2016.

[7] R. Sarno, Y. A. Effendi, and F. Haryadita, “Modified Time-Based Heuristics Miner for Parallel Business Processes,” *IRECOS (International Rev. Comput. Software)*, vol. 11, no. 3, pp. 249–260, 2016.

[8] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi, “Unsupervised discovery of intentional process models from event logs,” in *Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014*, 2014, pp. 282–291.

[9] G. Khodabandelou, C. Hug, R. Deneckere, and C. Salinesi, “Supervised intentional process models discovery using Hidden Markov models,” in *2013 IEEE Seventh International Conference on Research Challenges in Information Science*, 2013, p. 11 pp.

[10] X. Ding, T. Liu, J. Duan, and J.-Y. Nie, “Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network,” in *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI’15)*, 2015, pp. 2389–2395.

[11] S. Zhang and N. Wang, “Classification Model for Intent Mining in Personal Website Based on Support Vector Machine,” *Int. J. Database Theory Appl.*, vol. 9, no. 2, pp. 145–152, 2016.

[12] H. Zhou, J. Chen, G. Dong, H. Wang, and H. Yuan, “Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model,” *Mech. Syst. Signal Process.*, vol. 66–67, pp. 568–581, Jan. 2016.

[13] S. De Cnudde, J. Claes, and G. Poels, “Improving the quality of the Heuristics Miner in ProM 6.2,” *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7678–7690, 2014.

[14] O. J. K. P. J, “Penyimpangan Ketentuan Perbankan Yang Berindikasi Tindak Pidana Perbankan (Tipibank),” in *Seminar Deteksi Fraud dengan Audit Forensik dan Investigasi*, 2016.