Fraud Detection on Event Logs Using Fuzzy Association Rule Learning

Kadek Dwi Febriyanti[1], Riyanarto Sarno[2], Yutika Amelia Effendi[3]

[1]Department of Information Technology Management, [2,3]Department of Informatics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

e-mail: kadek.dwi.febriyanti16@mhs.mmt.its.ac.id, riyanarto@if.its.ac.id, yutika.effendi@gmail.com

Abstract—**Enterprise Resource Planning (ERP) is an enterprise information system designed to coordinate resources, information, and all activities required for a complete business process within a company. The most important requirement of an ERP system is integration. Integration in ERP system is used to combine the needs of a single software in a logical database, so that all departments in a company can share information and communicate with each other well. However, in its application, business process always changes dynamically. The changes will cause many variations to the original business process. This research discusses some frauds occured due to variations in business processes. Some frauds can be detected by using a method of process mining with Association Rule Learning approach. Process mining technique mine the fraud in the business process by checking the mismatch between event logs of business process with business process model of a company, which is better known as Standard Operating Procedure (SOP).**

Keywords—*association rule learning; fraud detection; fuzzy multi attribute decision making; fraud detection*

I. INTRODUCTION

Enterprise Resource Planning (ERP) has evolved as an integration tool, which is used to integrate all enterprise applications into a data storage center. ERP is easy to access in order to produce high efficiency for the company. However, in its application, business process always changes dynamically. The changes can occur due to several things, such as modification happened in the requirements or a policy in a business process that has been run before [1]. The changes will cause many variations to the original business process. In practice, variations of business process are sometimes executed correctly and incorrectly. The incorrect one may be contains some cheats, called fraud in terms of business process.

Fraud in business process can cause the losses for the company. There are many methods to detect the fraud. One of them is process mining which provides several advantages, i.e. conformance checking and control flow analysis in process mining can be used to compare actual data for original business process model [2]. This research discusses about fraud detection by using process mining with Fuzzy Association Rule approach. The evaluation of the proposed method uses case study of land management business process at sugar company in ERP system. First step is training process

to detect deviations which occur in the event logs generated by all activities in business process. Next step, the deviation data is analyzed by fuzzy association rule learning method to produce association rule and the confidence level. The deviation in this context is the violation behavior which is produced from the event logs of business processes and Standard Operating Procedure (SOP) of a company. In addition, this research is also explained about the calculation of fraud weight using fuzzy theory set multi attribute decision making.

The remainder of this paper is organized as follows. In Section II, we review some research which have been done in fraud detection. We present our proposed method in Section III. The results and analysis are explained in Section IV. Last, this paper is concluded with conclusions and sketched future works in Section V.

II. LITERATURE REVIEW

Section II gives a short summary of Fraud, Decision Point Analysis, Fuzzy Set Multi Attribute Decision Making (FMADM), and Association Rule Learning (ARL).

A. Fraud

Fraud is wrongful or criminal deception intended to result in financial or personal gain. Something or someone is categorized as a fraud if a person or thing intended to deceive others, typically by unjustifiably claiming or being credited with accomplishments or qualities [2]. In general, fraud occurs because there are three fundamental things happen together, namely incentives or the pressure of cheating, the opportunity to commit fraud, and attitude or rationalization to justify the act of cheating [3].

B. Decision Point Analysis

Definition of decision point is the moment a choice is made or must be made. In order to analyze the choices in a business process, we first need to identify those parts of the process model where the process is split into alternative branches, which is also called decision points. Based on data attributes associated to the cases in the event logs, we subsequently want to find rules for following one route or the others [4].

C. Fuzzy Set Multi Attribute Decision Making

Fuzzy Multiple Attribute Decision Making (FMADM) is a method used to find the optimal alternative of a number from many alternatives with certain criteria. The core of FMADM is to determine the weight value for each attribute, then proceed with the ranking process which will determine the alternatives already given before [5]. Basically, there are three approaches to find attribute weighted value, namely subjective approach, objective approach, and integration approach (combination of subjective and objective approach) [6].

Each approach has its advantages and disadvantages. In subjective approach, the weighted value is determined by the subjectivity of the decision makers, so that several factors in the alternative ranking process can be determined freely. Whereas in the objective approach, the weight value is calculated mathematically so that it ignores the subjectivity of the decision makers [7].

D. Association Rule Learning

Association Rule Learning (ARL) is one of the unsupervised data mining methods in which an item set is defined as a collection of one or more items [8]. Based on the behavior frequently found in a training dataset, the association rules are used for detection in a testing dataset. The dataset itself can imply the rules to change. In addition to its existing implementation [9], association rules is also suitable for generating filters against fraud. In fraud detection, the rules based on frequent data should be taken into account in which they enhance the system to detect a similar fraud in the following batch [10].

In ARL fuzzy, there is a difference in the calculation of the support value. For example the support for A => C in the fuzzy transaction set is supp (A U C) [12]. Suppose the value A = 0.4 and the value of C = 0.8, then the value of its support is min (0.4, 0.8) = 0.4. As for the calculation of the value of confidence equal to the concept of ARL [11].

III. THE PROPOSED METHOD

The basic idea proposed in this research is to present an automated solution for fraud detection based on learning in historical data. The association rules are generated from training data sets which the rules can show a degree of confidence to recognize whether a case is true or false.

A. Filtering Event Log

In this step, an analysis of the event logs is done to check whether the activity is intentionally bypassed or not in accordance with the business process model of a company. This analysis is performed using a modified Compliance Check, part of Conformance Checking analysis in ProM tool, so that it can show Skip Sequence and Skip Decision violations.

ProM tool is used to run the event logs and analyze the result. Fig.1 shows the discovered process model which has been analyzed using ProM.
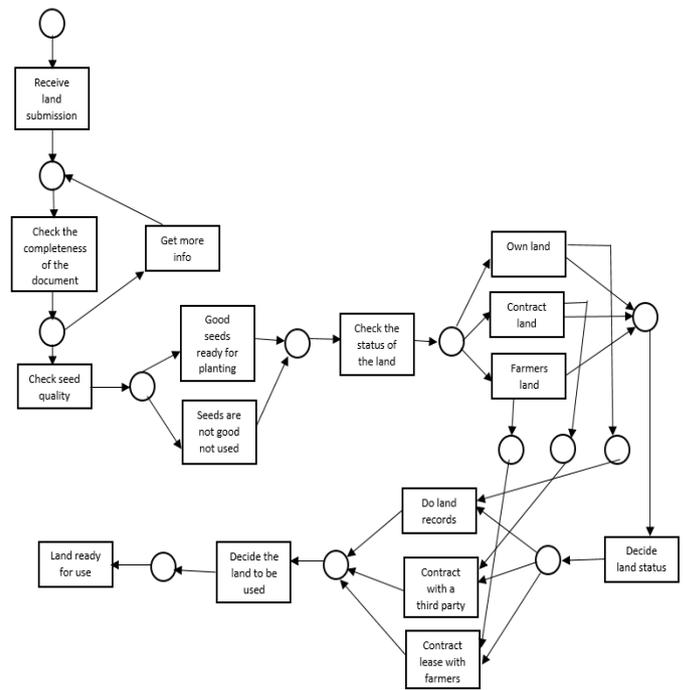

Fig.1 Discovered process model

B. Throughput Time Analysis

In this step, we analyze the time execution of all activities in the event logs by comparing them with time execution of Standard Operating Procedure (SOP) of a company. This is also done to detect violations in business process model.

C. Decision Point Analysis

In this step, we need to determine the decision point in the case study of land management business process at sugar company in ERP system. We also analyze the choices in a business process using ProM tool. Fig.2 and Fig.3 show that the process model has split into alternative branches.

Fig.2 shows that 'Location of Land' is divided into three branches. We can conclude that the process model has decision points.
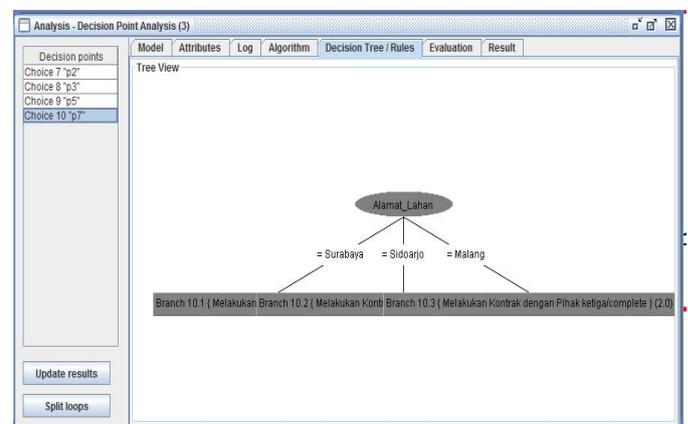

Fig.2 Decision Point Analysis using ProM tool for choice 10 in process model
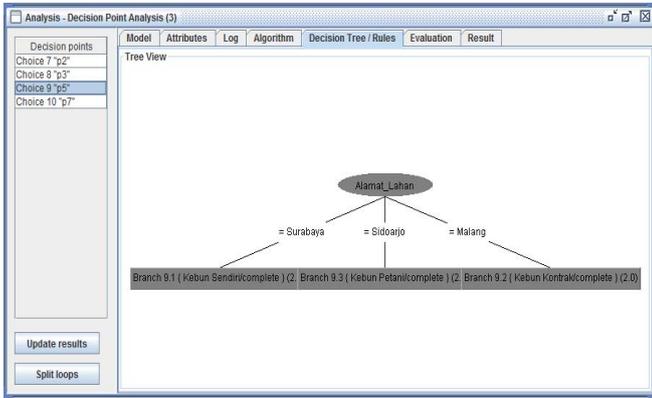
Fig.3 Decision Point Analysis using ProM tool for choice 9 in process model

D. Fuzzy Set Multi Attribute Decision Making

In this research, the data of expert interest assessment is taken from the research conducted by the experts. From the data, we calculate the weighted value, weighted value for each category, and attribute fraud. Weighted value is divided into four parts, i.e. lower bound weighted, middle weighted 1, middle weighted 2, and upper bound weighted. To calculate the four weighted values for each category, we can use (1), (2), (3), and (4).

$$Lower\ Bound = \frac{\sum_{k=1}^{n} a_k}{n} = \frac{0+0+0+0}{4} = 0 \qquad (1)$$

$$Middle\ Weight\ 1 = \frac{\sum_{k=1}^{n} b_k}{n} = \frac{0.3+0.3+0.3+0.3}{4} = 0.3 \qquad (2)$$

$$Middle\ Weight\ 2 = \frac{\sum_{k=1}^{n} c_k}{n} = \frac{0.3+0.3+0.3+0.3}{4} = 0.3 \qquad (3)$$

$$Upper\ Bound1 = \frac{\sum_{k=1}^{n} b_k}{n} = \frac{0.5+0.5+0.5+0.5}{4} = 0.5 \qquad (4)$$

Where n is the number of experts, the values $a, b, c,$ and d are the values of vectors a, vector b, vector c, and vector d in Table 1.

In the membership function formula, there is a value of x which is the probability value of the violation obtained from (5).

$$x = \frac{Number\ of\ violations}{Maximum\ Event\ Amounts\ that\ can\ be\ violated} \qquad (5)$$

The calculation of x is done for each attribute of violation. The result of this step will be used for the fuzzification phase to the membership class. Table I explains the level of importance, vector $a, b, c, d,$ and rating used in this research.

After that, a calculation is done to determine the weighted value for the violation data. Meanwhile, the violation data is taken from conformance checking using ProM tool. For the calculation violation data (P) is obtained by using (6).

$$P = \frac{\left(\frac{c_w+c_x+c_y+c_z}{4}\right)+\left(\frac{d_a+d_b+d_c+d_d}{4}\right)}{2} \qquad (6)$$

In (6), the notations of $C_w, C_x, C_y,$ and C_z represent the vector values w, x, y and z in Table II, which has correspond to the fuzzification value of the violation.

TABLE I. TABLE OF INTERESTS

Level of Importance	Vector				Rating
	a	b	c	d	
Very Important	0.7	1	1	1	100%-70%
Important	0.2	0.7	0.7	1	100%-50%
Fair	0.5	0.5	0.5	0.8	80%-20%
Weak	0	0.3	0.3	0.5	0%-50%
Very Weak	0	0	0	0.3	0%-50%

TABLE II. TABLE DEGREE OF VIOLATION

Attribute	w	x	y	z
Very Bad	0	0	0.1	0.2
Bad	0	0.1	0.2	0.3
Fair	0.1	0.2	0.3	0.4
Good	0.2	0.3	0.4	0.5
Very Good	0.3	0.4	0.5	0.6

In addition, the notations of $D_a, D_b, D_c,$ and D_d represent the vector values a, b, c and d in Table I which also has correspond to the importance value given by the expert. Next step, the value of P is categorized into the class of violation according to the membership function of the degree of violation.

After the value of the violation has been adjusted into the value of the interest of the experts, the next step is to calculate the weight for the violation by using the same equation when calculating the weight of interest with the value of the notation n = number of +1 experts.

Next, for the weight of interest and violation, the final rating calculation is used to calculate lower bound weighted, middle weighted 1, middle weighted 2, and upper bound weighted using (7).

$$Final\ Rating = \frac{1}{x}\ [(s_{CI} \times W_{CI})+ [(s_{Cn} \times W_{Cn}) \qquad (7)$$

After the four final weighted ratings have been calculated, the weighted of fraud of a case is the sum of the four final weighted ratings. The calculation of weighted fraud is done for each process to be examined using (8). The value of weighted of fraud is used to determine the association rule using Fuzzy ARL.

$$Fraud\ Weight = Fr(LB) + Fr(M1) + Fr(M2) + Fr(UB) \qquad (8)$$

E. Fuzzy Association Rule Learning

This method is used to determine the association rules of violations which occur in business process. The data which will be processed is data violation with weighted of fraud for each process. The probability value for each attribute violation

TABLE III. FRAUD DETECTION WITH ASSSOCIATION RULES

Case	Rules	Confidence	Weight	Class
1	Nothing	0	0	No Fraud
2	Nothing	0	0	No Fraud
3	wDuty DecL-Fraud	0.27878	0.0049	No Fraud
4	wDecisionL-Fraud	0.4987	0.1226	No Fraud

- is then used to calculate the degree of membership. To determine the association rule, each attribute of violation is divided into three types, based on the number of violations committed, i.e. Low, Middle, and High.

The next step is to calculate the degree of membership of each attribute violation against three types or class membership low, middle, and high. The items which pass the selection are then combined with other items to form a new itemset. This process is run continuously until no more formation of the item can be formed or it can be limited to the formation of the itemset until itemset. In this method, rules are defined for fraud cases and for non-fraud cases. The determination of this rule is necessary to correctly define both fraud and non-fraud cases.

F. Detecting Fraud with Association Rule Data

This method is used as a testing phase to detect fraud in business processes by matching attribute violations with the result of association rules. If a case contains an association rule which is not defined as fraud, then the case is directly defined and categorized as a case which is not fraud, despite having a high confidence value. Whereas if a case contains fraud association rules, then the case is categorized as fraud. Table III shows the result of fraud detection.
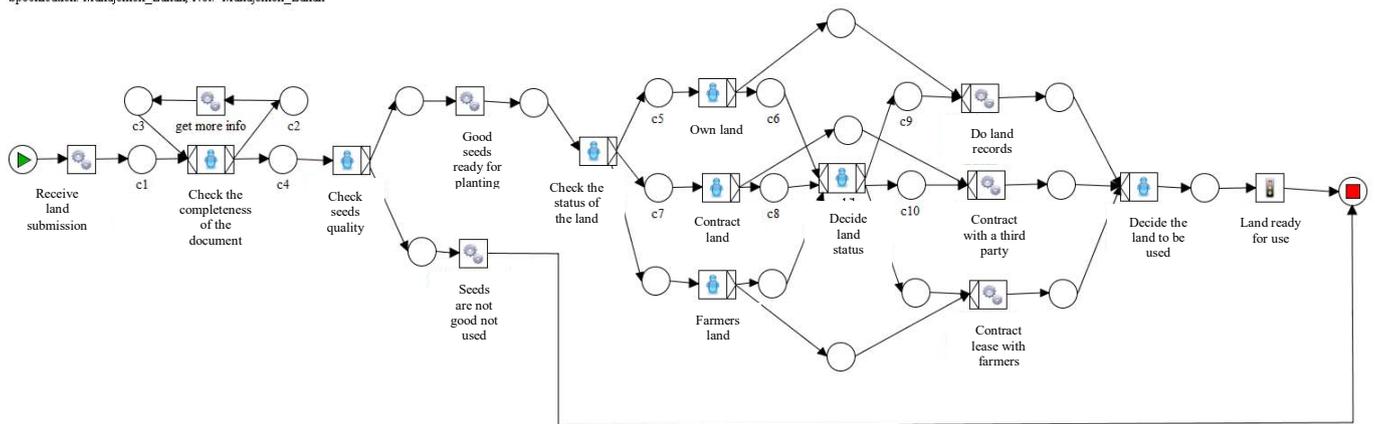
IV. EVALUATION

In this research, land management applications for sugar companies has been implemented as case studies. Some artificial event logs which contain cheats or fraud have been created. Discovered process model using ProM is shown in Fig 1. Fig.1 illustrates the flow control model of Standard Operating Procedure (SOP) of sugar cane management application. The description of the business process in this research represents using YAWL (*Yet Another Workflow Language*) editor. YAWL is an application used to create and simulate workflows of a business system, using a compact and powerful modeling language. YAWL can also create a log file of a business process simulation, which we can do process mining against it. This workflow design starts from 'Receive land submission documents'. First of all, we check the completeness of the document. The next processes are checking the seeds quality and check the status of the land. 'Check the status of the land' is done to know which land is used for the process of planting sugarcane. After we find the decisions which land will be used, then the land is ready to be planted sugarcane and land management process is over.

The violations that occurred will be evaluated in the experiment using two cases. The first case study is the business process of land management applications in sugar companies. Further tests are performed using data testing with additional rules and accuracy. For the second case study is the process of 'Check the status of the land' in ERP system. After we do testing the two cases, we compare the results of the two cases.
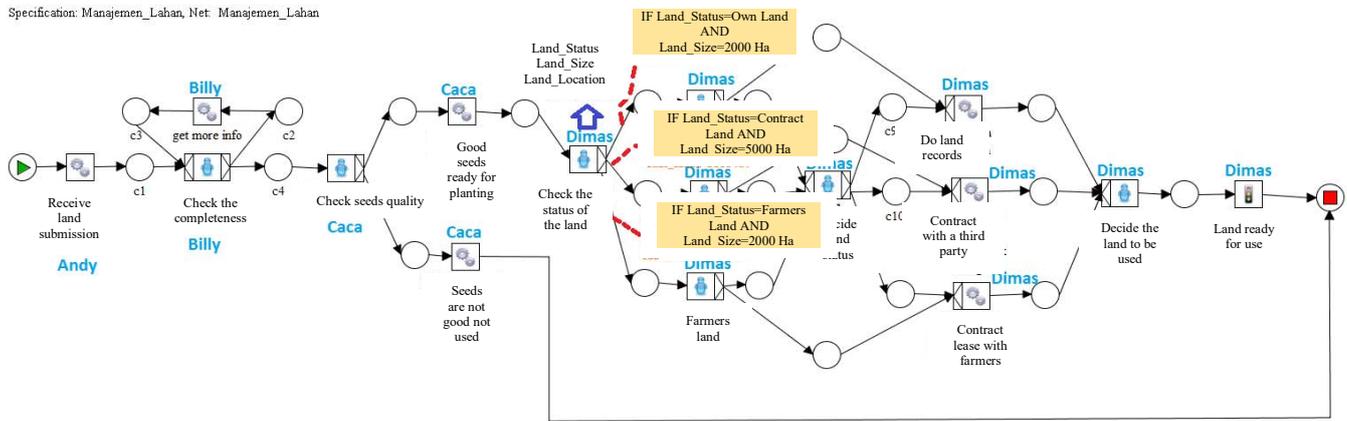
Specification: Manajemen_Lahan, Net: Manajemen_Lahan

Fig.4 Business Process Sugarcane Farm Management

Fig.5 Misrepresentation of Resource

V. EVALUATION AND TESTING

In this research, there are two cases which will be implemented. The first case study is the business process of land management applications in sugar companies. Variations of violations which may occur in this case are 150. The experiment was conducted using 100 cases, consisting of 75 cases for training and 25 cases for testing. For 35 cases training, consisting of 5 cases which are not fraud without violation at all (weight = 0), 5 cases which are not fraud and violation is only one type with low intensity, and 25 cases which are fraud. After we analyze the business process, we get the results Table IV.

TABLE IV. ACCURACY METHOD WITHOUT CASE RULE

Min Confidence	TP	FP	TN	FN	Accuracy
0.1	67	15	12	0	0.75
0.2	67	15	12	2	0.74
0.3	65	15	12	2	0.46
0.4	65	13	13	3	0.43
0.5	62	12	10	3	0.59
0.6	54	10	10	15	0.61

TABLE V. ACCURACY METHOD WITH CASE RULE

Min Confidence	TP	FP	TN	FN	Accuracy
0.1	70	7	30	0	0,093
0.2	75	7	30	0	0,096
0.3	67	6	27	25	0,089
0.4	65	6	28	20	0,091
0.5	66	6	25	20	0,094
0.6	60	6	30	20	0,086

The measurement of this accuracy is determined using the minimum confidence value. For accuracy calculation results, we use (9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

The accuracy calculation in the first case study is the highest accuracy of 73% for the minimum values of confidence 0.1 and 0.2. Table IV shows that there is a large number of false positive with probability of 17/35, whereas the number of false negative is less with probability 0. Based on this result, we can conclude that the normal case is not properly defined yet. Thus, it is necessary to establish a rule which defines a normal case to reduce the number of false positive and increase the accuracy.

Furthermore, the experiments were performed in data set with additional rules and the accuracy increased to 96% with the probability that the number of false positive reduced to 4/35. The results of the calculation of accuracy with additional rules can be seen in Table V.

After the training and testing were performed and we obtain the data, the accuracy of the data testing calculation is the same for the first case study. The accuracy calculation was performed in one condition, i.e. conditions with the same rules for the first case study. This is done to determine whether the established rules can be used to detect fraud in the land management business process. Table VI shows the final result of the accuracy calculation results.

TABLE VI. ACCURACY AGAINST DATA TESTING ON ERP CASE STUDY "CHECK THE STATUS OF THE LAND"

Min Confidence	TP	FP	TN	FN	Accuracy
0.1	39	5	37	0	0,097
0.2	39	2	50	0	0,096
0.3	39	2	27	1	0,089
0.4	38	2	28	1	0,091
0.5	38	2	25	9	0,094
0.6	38	2	30	9	0,086

VI. CONCLUSION

This research paper discusses some frauds occured due to variations in business processes. Some frauds can be detected by using a method of process mining with Association Rule Learning approach. Process mining technique can mine the fraud in the business process by checking the mismatch between event logs of business process with business process model of a company, which is better known as Standard Operating Procedure (SOP). The conclusions from the experiment which have been done before are:

- From the results of the experiments which have been conducted in two cases, conformance checking method in process mining can be used to detect deviation in business process.

- Fuzzy association rule method can be used to determine fraud based on business process. This is indicated by the accuracy value obtained from the fraud detection process for both cases that reached more than 90%.

- In addition, the establishment of specific additional rules for handling non-fraud cases can help to improve accuracy as it reduces the occurrence of False Positive in fraud detection, which is shown in Table V. The accuracy to detect fraud in the first case study without using the additional rule is only 74%. And in Table VI, when we use an additional rule, the accuracy to detect fraud increases to 96%.

ACKNOWLEDGMENT

REFERENCES

[1] S. Gupta, W. M. P van der Aalst, A. J. M. M.Weijters, and A. K. Alves de Medeiros, Workflow and Process Mining in Healthcare, Eindhoven, 2007.

[2] S. Huda, R. Sarno, T. Ahmad, and H. A. Santoso, "Identification of Process-based Fraud Patterns in Credit Application," in 2014 2nd International Conference on Information and Communication Technology (ICoICT), 2014, pp. 84–89.
https://doi.org/10.1109/ICoICT.2014.6914045

[3] R. Sarno, R. D. Dewandono, T. Ahmad, M. F. Naufal, F. Sinaga, Hybrid Association Rule Learning and Process Mining for Fraud Detection, IAENG International Jounal of Computer Science, 42 (2), pp. 59-72, 2015.

[4] A. A. Pranatha, A. H. Nur Ali, and Sholiq, Analisis Perikiraan Biaya Pembuatan Enterprise Resource Planing (ERP) Modul Pabrik Gula Diperusahaan Perkebunan Dengan Metode Cocomo II, Surabaya, 2012.

[5] M. P. Barreiros, A. Grillo, V. Cruz Machado, and M. R. Cabrita, "Applying Fuzzy sets for ERP Systems Selection within the Construction Industry," 2010.

[6] Hermawan and R. Sarno, "Developing Distributed System with Service Resource Oriented Architecture," TELKOMNIKA (Telecommunication, Computing, Electronics and Control), vol. 10, no. 2, pp. 389–399, 2012.
https://doi.org/10.12928/telkomnika.v10i2.815

[7] C. F. Lee and C. H. Tsai, Efficient Associating Mining Approaches for Compressing Incrementally Updatable Native XML Databases, IAENG International Journal of Computer Science 33 (1) (2007), pp. 78-85.

[8] S. C. Abou and T. M. Dao, Association Rules Mining Approach to Mineral Processing Control, Engineering Letters, 18 (2) (2010), pp. 156-164.Report to the Nations on Occupational Fraud and Abuse," ACFE, 2012.

[9] R. Sarno, E. W. Pamungkas, D. Sunaryono, and Sarwosri, "Business Process Composition Based on Meta Models," in 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2015, pp. 315–318.
https://doi.org/10.1109%2Fisitia.2015.7219998J.

[10] J. J. Stoop, "A case study on the theoretical and practical value of using process mining for the detection of fraudulent behavior in the procurement process," in *Process Mining and Fraud Detection*, Netherlands, Twente University, 2012.

[11] M. Delgado, "Mining Fuzzy Association Rules: An Overview," in *BISC Conferece*, 2003.