Decision Tree and Fuzzy Logic in The Audit of Information System for Tax Letter Issuance

Irfanur Ilham Febriansyah
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya

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

Ratih NE Anggraini
Informatics Department
Institut Teknologi Sepuluh Nopember
Surabaya
ratih_nea@if.its.ac.id

2022 International Electronics Symposium (IES) | 978-1-6654-8971-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IES55876.2022.9888372

***Abstract*— An information system audit is a way that can be taken to ensure the quality of information systems in an organization. One of the problems faced is objectivity in assessing quality which can be overcome by mining the processes, where the audit process utilizes event logs from the audited information system. In the domain of government in Indonesia, information system audit has not been widely conducted, as has never been done in the information system for tax letters issuance at the Regional Revenue Agency of East Java Province. An audit is conducted to determine whether the information system has been running in accordance with the needs of the organization. This study implements part of the process mining: the system conformance assessment using Fuzzy. The dataset used as input comes from the event log classification process using the Decision Tree classifier. The study results indicate that the system alignment score is 2.754031 or medium predicate.**

Keywords*— *Information System Audit, Process Mining, Tax Letters Issuance, Decision Tree, Variable Correlation, Fuzzy

I. INTRODUCTION

An information system audit examines the organization's information system to determine whether the related activities achieve the organization's objectives [1]. One aspect of the audit is integrity, meaning the auditing party must assess the system as formally, independently, and objectively as possible. Thus, the value obtained describes the quality of the audited system. For that aspect to be fulfilled, assessment-based process mining can be used, allowing the system to assess itself based on existing data.

Process mining is a discipline that bridges data-centric analysis techniques (data science) with traditional model-based process analysis (process science) [2]. The goal is to provide fact-based insights and to support process improvements. As part of data science, process mining seeks to confront event data and process models. In contrast, process mining is discovering process models from event data as part of process science. Therefore, by using event data generated by an information system, process mining can be used for information system audit purposes, both from data science and process science.

Assessments in information system audits are usually done using specific frameworks. One of the frameworks is Control Objective for Information Technologies (COBIT). Five domains manifest COBIT, namely Evaluate, Direct, and Monitor (EDM); Align, Plan, and Organize (APO); Build, Acquire, and Implement (BAI); Deliver, Service, and Support (DSS); Monitor, Evaluate, and Asses (MAE) [3].

The audit's implementation is determined on the BAI domain to limit the study's scope. This domain aims to provide solutions and deliver them in service. IT solutions must be identifiable, developed, acquired, implemented, and integrated into business processes to realize an IT strategy. This domain also covers changes in the maintenance process of existing systems to ensure that IT solutions can continue to meet business objectives.

To further narrow the study, the bai domain goal is determined by taking the sub-goal BAI01.01: Maintain a standard approach for program management. This goal can be translated as follows: Maintain a standard approach for program management that enables governance and management review, decision-making and delivery management activities. These activities should focus consistently on business value and goals (i.e., requirements, risk, costs, schedule, and quality targets).

Based on the goal, the focus of the study is to know the performance of the system, whether it has become a solution in accordance with the business processes desired by stakeholders, as well as knowing the level of maturity of information system performance to be given recommendations related to system improvement or development.

In the realm of government in Indonesia, especially in East Java province, information system audit has not been widely conducted. The study seeks to assess the maturity of information systems using mining process techniques on several datasets generated by the system, then infers the system's maturity using fuzzy logic, which has not been implemented on information system audit research.

II. RELATED WORKS

Studies of information system audits have been widely conducted, mostly reviewing the assessment of the maturity or capabilities of the system using specific frameworks. One of the frameworks used is COBIT, demonstrated in [4-9]. Another framework is Information Technology Infrastructure Library (ITIL) used in [10-14].

Other evaluation standards can also be used for information system audits, as conducted by Wulandari in his study, which uses ISO 25010 standards to conduct internal quality audits on information systems [15]. A combination of frameworks was used in other studies, such as those conducted by Andry, who evaluated information systems based on COBIT 5 and ITIL V3 [16].

The studies describe how an information system is audited using one or more domains on the COBIT framework, ITIL, other evaluation standards, or combinations with other frameworks. The studies are also accompanied by results and recommendations provided to information system managers. Research is conducted using reference studies, field

observations, and interviews with relevant stakeholders in the audited system.

Other research related to information system auditing is in mining processes. Existing process mining research seeks to explore and modify the mining process in information system audits in an enormous scope, as done by Radityohutomo et al., who have implemented genetic algorithms in process mining [17]. Roubtsova et al. have combined process mining and participatory workshop, which contains IT Expert, Business Expert, dan Audit Expert [18]. Erdogan et al. have developed a goal-driven evaluation method on process mining [19], and Jokonowo et al. have performed the extraction process based on event log [20].

III. Research Method

The process mining is conducted by focusing on the conformance assessment section without modelling the business processes within the system. For auditing information systems, data that corresponds to certain parameters combined with fuzzy logic to, in conclusion, the system's value (Fig.1).

A. Audited Information System

The case study is on the information system of issuing tax papers in the East Java Regional Revenue Agency (Bapenda Jatim). This system is used to be a medium of management in issuing tax letters that are the authority of Bapenda Jatim, especially for Motor Vehicle Tax (PKB).

There are three types of tax letters issued by this system in accordance with the type of taxpayer noncompliance in Fig.2, namely The Letter of Collection of Objects and Subjects (SPOS), Tax Calculation Note (NPP), and Tax Bill Note (NTP). SPOS is issued to taxpayers with non-compliant category 2, NPP is issued to taxpayers with non-compliant category 3, NTP is issued to taxpayers with non-compliant category 3.

B. Assessment Parameters

Before determining the assessment parameters, the intersection of performance indicators is identified. Based on the results of identification (see Fig.3), it was found that the information system has a good maturity value if the system performance indicators and indicators of organizational performance are met simultaneously, namely the issuance of tax letters in accordance with the prediction of taxpayer noncompliance as well as increasing taxpayer compliance.

The two indicators are then derived into assessment parameters, namely by comparing: (1) tax letter issuance data with a prediction of taxpayer noncompliance; (2) tax letter issuance data with Motor Vehicle Tax payment data.

If the data of the issuance of tax letters is directly proportional to the prediction of taxpayer noncompliance, indicating that the information system has carried out its primary function, which is to issue tax letters. On the other hand, if the tax letter issuance data is directly proportional to the Motor Vehicle Tax payment data, it indicates that the information system meets performance indicators raising the taxpayer's compliance rate.

C. Dataset

To be able to calculate the assessment parameters, three datasets derived from Bapenda Jatim are used, namely Tax Letter Issuance Data (Table 1), Motor Vehicle Tax Payment Data (Table 2), and taxpayer noncompliance prediction data.

The two data first were obtained by querying the raw material data, while the latter was obtained from the classification process. The data used is limited to the Wonorejo area of Rungkut subdistrict, Surabaya city, with tax expiration in 2017-2020.

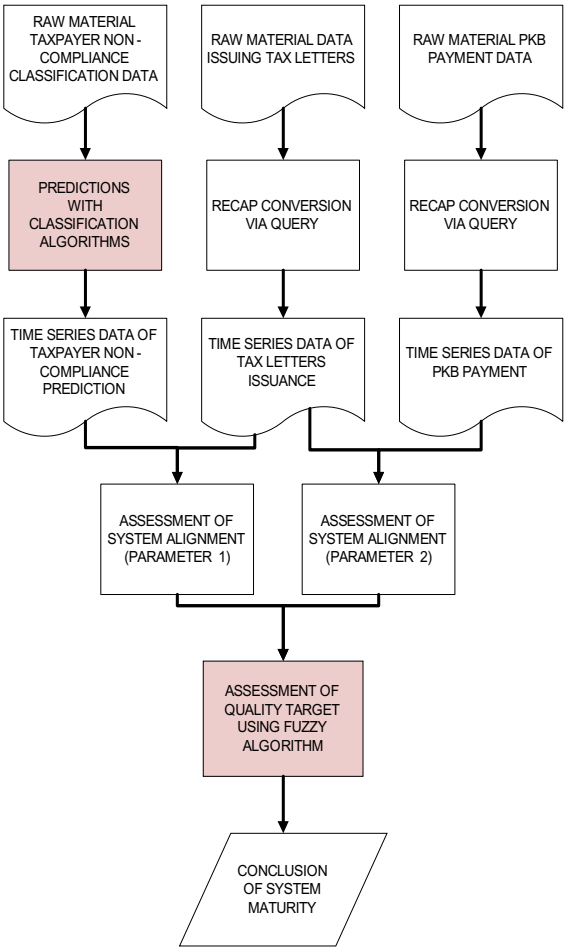


Fig.1. Proposed Method

Fig.2. Category of Taxpayer Non-Compliance

Fig.3. Intersections of Organizational Performance Indicator and Tax Letter Issuance Information Systems Performance Indicator

Table 1. Tax Letter Issuance Data

Validity Period	SPOS	NPP	NTP
2017-01-01	180	145	119
2017-02-01	239	214	179
2017-03-01	198	175	157
2017-04-01	195	164	136
2017-05-01	236	205	186

Table 2. Motor Vehicle Tax Payment Data

Validity Period	Number of Payment
2017-01-01	826
2017-02-01	925
2017-03-01	912
2017-04-01	1011
2017-05-01	1111

D. *Decision Tree for Taxpayer Non-Compliance Classification*

The classification of taxpayer noncompliance is done with the Decision Tree classifier. This classifier is a popular supervised learning method because of its ease of understanding and implementation. Decision trees have a hierarchical structure like an inverted tree consisting of root nodes, leaf nodes, and branches. The root node is the top that shows the most dominant feature, the leaf node is the impacting feature underneath, and the branch indicates the decision.

The input dataset used is raw material data on the payment of Motor Vehicle Taxes. A classifier with an accuracy above 0.95 will classify taxpayer noncompliance on the raw dataset of motor vehicle tax payment material. The results of the classification are then recapitulated into time-series data.
Classifier accuracy is obtained by calculating the ratio of the number of correctly classified data instances over the total number of data instances based on the results of classification visualized using a confusion matrix. Accuracy calculations are shown in Equation 1, where TN is True Negative, TP is True Positive, FN is False Negative, and FP is False Positive.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (1)$$

E. *Fuzzy Logic for System Quality Assessment*

System alignment assessment aims to see how well a system performs tasks in accordance with system assessment parameters. Two parameters of assessment that have been determined are assessed.

Assessment of the quality of the system is done by calculating the correlation between 2 variables related to the parameters: (1) tax letter issuance data (Table 1) with taxpayer noncompliance prediction data generated in the taxpayer's noncompliance classification process; and (2) tax letter issuance data (Table 1) with Motor Vehicle Tax payment data (Table 2).

The correlation value indicates how strong the relationship between the two variables is. The range of correlation values is between -1 and 1. The closer the correlation value is to 0, the weaker the relationship between the two variables. A positive correlation, in which the correlation value is between 0 and 1, indicates that the increase in one variable will also increase its dependent variable. On the other hand, there can be a negative correlation, where the correlation value is between 0 and -1, indicating that the increase in one variable will decrease another.

The correlation value obtained is then processed using Fuzzy logic to determine the Alignment System Score and predicate associated with the value. Fuzzy is defined as vague, which means a value can be of right or wrong value simultaneously. Fuzzy is expressed in degrees of membership and degrees of truth. Therefore, something can be said to be partly true and partly wrong simultaneously. The truth and error values depend on the weight of its membership.

The terms rule, antecedent, and consequent are known in fuzzy logic. The rule in fuzzy is a set of IF-THEN statements that store the practical knowledge of human operators about the process [21]. The antecedents are the premises of the rule (IF section) that must be true before the conclusions of the rule (THEN section) can be asserted. The consequents are the conclusions of the rule.

In this study, the antecedent used is six pieces of correlation values between variables in the assessment parameter, which has a range of -1 and 1, with membership functions of 3 pieces, namely Poor, Average, and Good. Visuals of antecedent are shown in Fig.4. Consequently, this rule is in the form of an alignment score of the audited information system, with membership functions as many as three pieces, namely Low, Medium, and High, as seen in Fig.5.

The rule is determined as many as six items, where rule id 1-3 is a rule that uses input correlation parameter assessment 1, while rule id 4-6 is a rule that uses input correlation parameter assessment 2, which is shown in Table 3.

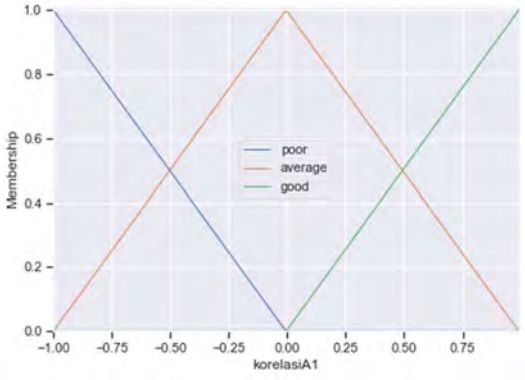


Fig.4. Membership Function of Antecedent

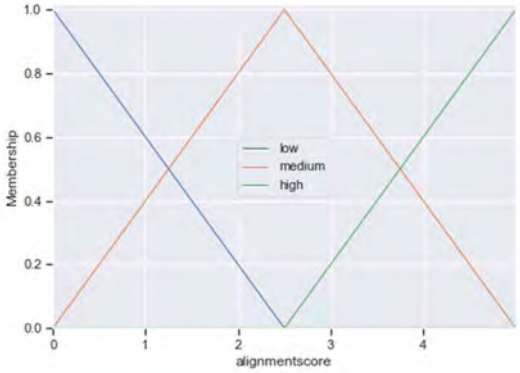


Fig.5. Membership Function of Consequent

Table 3. Rule Fuzzy used

Rule ID	A1	A2	A3	C	P
1	Poor	Poor	Poor	Low	1
2	Good	Good	Good	High	1
3	Good	Average	Poor	Medium	1
4	Poor	Poor	Poor	Low	2
5	Good	Good	Good	High	2
6	Good	Average	Poor	Medium	2

A1: Antecedent 1; A2: Antecedent 2; A3: Antecedent 3; C: Consequent; P: Parameter Id

IV. Experiment and Result

A. Experiment Tool

The experiments are performed using Python version 3.8.5, Scikit-Learn library version 0.23.2, and Scikit-Fuzzy version 0.2.

B. Classification of Taxpayer Non-compliance

The raw material dataset of Motor Vehicle Tax payment material is available, amounting to 52924 data, with a total of 18 features and without a target label. For the dataset used as a class input to have a category label, first does the target label manually by Bapenda East Java on data amounting to 5000 rows.

Before conducting the training and testing process with a classifier, the data is prepared in advance by ensuring all data is numerical type. It is converted into strings for date-type data with the format 'YYYYMMDD', then converted to numeric. At the same time, categorical data is converted to numeric using the One Hot Encoding method, the number of features in the dataset changes to 49 features and one target label. The data is then divided into trains and test sets with test sets amounting to 20% of the total 5000 data and then normalized to be used as input on the classifier.

Experiments on the test set found a confusion matrix, as seen in Fig.6, which showed that the misclassification of data was very small, characterized by black blocks, which amounted to 13 out of 1000 rows. On the other hand, correctly classified data amounting to 987, so the accuracy value of 987 / 1000 = 0.987. Therefore, we decided to use the model to classify the raw dataset of motor vehicle tax payment material.

After the classification process, taxpayer noncompliance data is obtained with the distribution of labels, as shown in Fig.7. The data is then carried out to make it a dataset time series (Table 4).

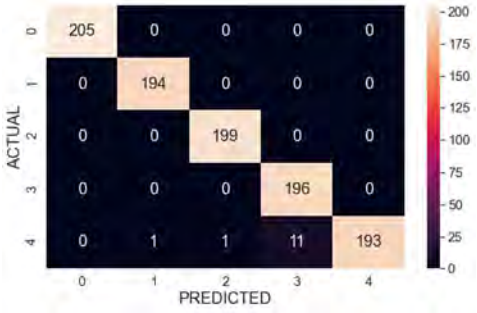


Fig.6. Confusion Matrix of Classification Result

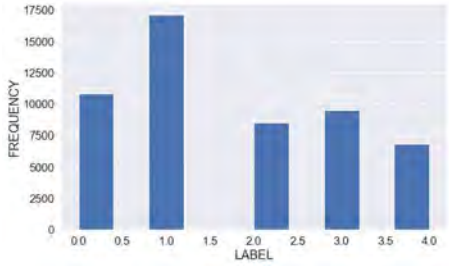


Fig.7. Label Distribution of Taxpayer Non-compliance Classification Results

Table 4. Taxpayer's Non-compliance Prediction Data

Validity Period	Non-Compliant Category	Number of Non-Compliant Taxpayer
201701	0	853
201702	1	23
201703	4	51
201704	0	850
201705	1	41

C. Parameters Fulfillment Assessment

In the assessment of achievement in parameter 1, namely the comparison of tax letter issuance data (Table 1) with data on the prediction of taxpayer noncompliance (Table 4) is done comparing two variables that have relationships based on the nature of noncompliance, namely: (1) the issuance of SPOS tax letters with predictions of noncompliance category 2; (2) issuance of NPP tax letter with the prediction of category three noncompliance; and (3) the issuance of NTP tax letters with predictions of category four noncompliance.

For parameter 2, namely the comparison of tax letter issuance data (Table 1) with Motor Vehicle Tax payment data (Table 2), the variable comparison is made based on the nature of its noncompliance with PKB payment variables.

Table 5. Comparison of Variables in Parameter Assessment 1

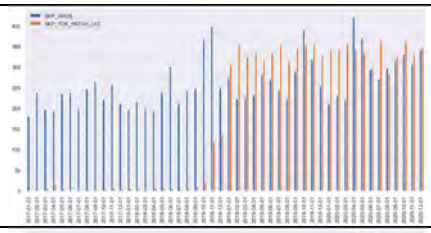
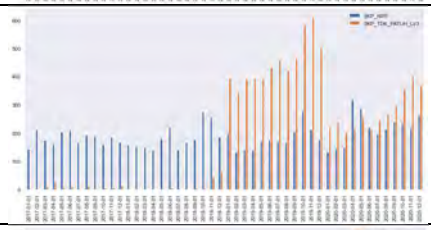
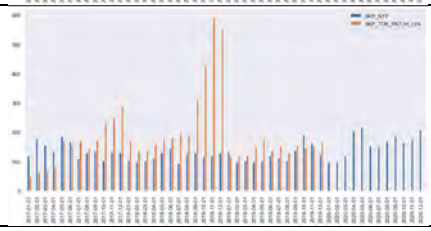
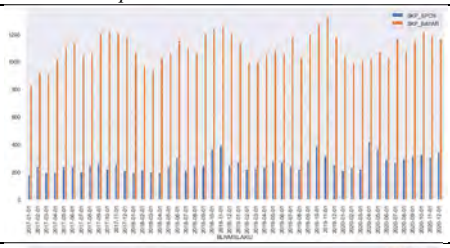
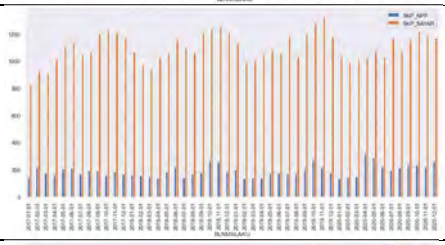
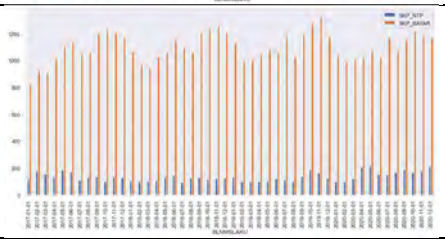
Correlation Id	*Variable comparison visualization*	*Annotations*	*Correlation Score*
A1		*The variable compared is the issuance of SPOS tax letters with predictions of category 2 noncompliance.*	*0.418551*
A2		*The variable compared is the issuance of NPP tax letters with predictions of category 3 noncompliance.*	*0.125239*
A3		*The variable compared is the issuance of NPP tax letters with predictions of category 4 noncompliance.*	*-0.106176*

Table 6. Comparison of Variables in Parameter Assessment 2

Correlation Id	*Variable comparison visualization*	*Annotations*	*Correlation Score*
B1		*The variable compared is the issuance of SPOS tax letters with PKB payment variables.*	*0.542947*
B2		*The variable compared is the issuance of NPP tax letters with PKB payment variables.*	*0.445595*
B3		*The variable compared is the issuance of NTP tax letters with PKB payment variables.*	*0.233519*

Tables 5 and 6 show visualization and correlation values of comparing two variables related to the assessment of parameters 1 and 2. The correlation value is further used as an input in the system's quality assessment using Fuzzy logic.

Assessment of system alignment is only done on data with noncompliance classifications 2, 3, and 4. This is done because the classification can be assessed using available data to check the correlation of variables, while the classification of noncompliance 0 and 1 is not assessed.

D. System Quality Assessment

System quality assessment is done with Fuzzy logic. Computation of the alignment of the scoring system is carried out based on antecedent, consequent, and specified rules. The input used is the correlation values A1, A2, A3, B1, B2, and B3 that has been obtained from the previous process and found a system alignment score of 2.754031 or the predicate of the medium, shown in Fig 8.

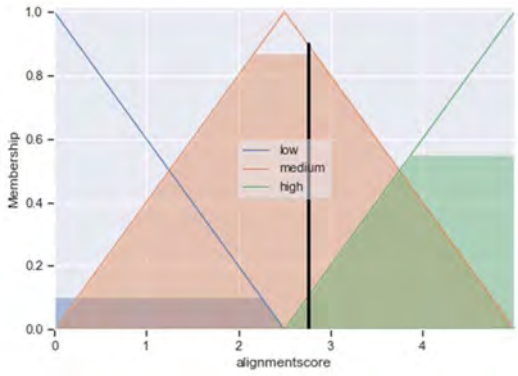


Fig.8. Visualization of System Quality Assessment

V. CONCLUSIONS

Correlation values on related variables that have been specified on the assessment parameters are used for the alignment assessment of the system. Variables are taken from datasets obtained from the recording by the information system assessed. Variable correlation values are further used as inputs in Fuzzy logic to produce outputs in the form of system alignment score and system quality assessment predicate.

The assessment results were obtained at 2.754031 or medium predicate. It can be said that the system has not been maximal in meeting system performance indicators as well as organizational performance indicators simultaneously. Therefore, stakeholders need to evaluate and adjust to the system's performance.

For the development of subsequent research, it is recommended to add system assessment parameters according to the complexity of the assessed system. With the increase in the assessment parameters, it is also necessary to be supported by the addition of the number and type of logs from the system. Other developments can be implementations of other domains from audit frameworks.

REFERENCES

[1] A. R. Otero, *Information Technology Control and Audit*, no. July 2018. 2018.

[2] W. Van der Aalst, *Process mining: Data science in action*. 2016.

[3] ISACA. and J. W. Lainhart, *COBIT 5: A business framework for the governance and management of enterprise IT COBIT 5*, vol. 34, no. 1. 2012.

[4] B. H. Mesa and J. F. Andry, "Evaluation of Information System Effectiveness Level Using COBIT Framework 5," *CogITo Smart J.*, vol. 4, no. 1, p. 148, 2018, doi: 10.31154/cogito.v4i1.89.148-159.

[5] A. A. G. O. K. Adnyana, I. N. D. Kotama, and I. B. A. Swamardika, "Audit Graduation Information System (SIMUDAPAPI) in Udayana University using COBIT 5 in Deliver, Service, and Support Domain," *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 2, 2018.

[6] J. F. Andry and A. Chakir, "Assessment It Governance of Human Resources Information System Using Cobit 5," *Int. J. Open Inf. Technol.*, vol. 8, no. 4, pp. 59–63, 2020.

[7] B. A. Bahari, F. Adnan, and B. Prasetyo, "Audit Capability Level Using COBIT 5.0: A University Customer Care Center at University of Jember," *Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019*, pp. 5–12, 2019, doi: 10.1109/ICOMITEE.2019.8921118.

[8] Fathoni, N. Simbolon, and D. Yunika Hardiyanti, "Security audit on loan debit network corporation system using cobit 5 and iso 27001: 2013," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012033.

[9] G. G. Prapenan and G. C. Pamuji, "Information System Security Analysis of XYZ Company Using COBIT 5 Framework and ISO 27001:2013," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 879, no. 1, 2020, doi: 10.1088/1757-899X/879/1/012047.

[10] H. W. Kim, B. R. Kang, D. S. Kim, and S. C. Moon, "A Design of University Information System Operatio Audit Model based on ITIL V3," *J. Inf. ...*, vol. 25, no. September, pp. 29–41, 2018, [Online]. Available: http://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE07543805.

[11] N. L. A. Dewi and M. Sudarma, "IT Audit with ITIL on Business Process , Application and IT Infrastructure in Bali Rattan Bag," vol. 6, no. 2, pp. 61–64, 2021, [Online]. Available: http://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE07543805.

[12] K. Christianto and J. F. Andry, "Evaluating Maturity Level Using Framework ITIL: A Case Study of Service Desk's," *Int. J. Inf. Technol. Bus.*, vol. 1, no. 1, pp. 16–23, 2018, doi: 10.24246/ijiteb.112018.16-23.

[13] A. Hermanto and G. Kusnanto, "Evaluation of the information technology system services for medium higher education based on ITIL (A case study of polytechnic XYZ)," *Proc. 2017 4th Int. Conf. Comput. Appl. Inf. Process. Technol. CAIPT 2017*, vol. 2018-Janua, pp. 1–7, 2018, doi: 10.1109/CAIPT.2017.8320678.

[14] M. Ilham, A. Eliyana, T. Handriana, I. Usman, and E. Mufidah, "DigitalCommons @ University of Nebraska - Lincoln The Effect of Information Technology Audit For E-Health of Indonesia Using ITIL Framework V . 3 Domain Service Design The Effect of Information Technology Audit For E-Health of Indonesia," 2020.

[15] D. A. P. Wulandari and M. D. W. Aristana, "Analysis Evaluation Management Information System Audit Internal Quality," *J. Electr. Electron. Informatics*, vol. 5, no. 1, p. 5, 2021, doi: 10.24843/jeei.2021.v05.i01.p02.

[16] J. Fernandes Andry, F. Sakti Lee, and L. Liliana, "Evaluation of The Human Resource Information System With COBIT 5 and ITIL V3 (Case Study:Pharmaceutical Company)," *Int. J. Eng. Inf. Syst.*, vol. 5, no. May, pp. 123–129, 2021, [Online]. Available: www.ijeais.org/ijeais.

[17] Y. Radityohutomo, G. A. A. Wisudiawan, A. Alamsyah, and A. Herdiani, "Implementation of Genetic Process Mining to Support Information System Audit," *8th Int. Conf. Sustain. Collab. Business, Technol. Inf. Innov.*, vol. 01, pp. 1–7, 2018, [Online]. Available: http://libraryeproceeding.telkomuniversity.ac.id/index.php/scbtii/article/download/5500/5481.

[18] E. Roubtsova and N. Wiersma, "A practical extension of frameworks for auditing with process mining," *ENASE 2018 - Proc. 13th Int. Conf. Eval. Nov. Approaches to Softw. Eng.*, vol. 2018-March, no. Enase 2018, pp. 406–415, 2018, doi: 10.5220/0006798904060415.

[19] T. G. Erdogan and A. Tarhan, "A goal-driven evaluation method based on process mining for healthcare processes," *Appl. Sci.*, vol. 8, no. 6, 2018, doi: 10.3390/app8060894.

[20] B. Jokonowo, R. Sarno, and S. Rochimah, "Extracting audit trail data of port container terminal for process mining," *Proc. 2019 Int. Conf. Inf. Commun. Technol. Syst. ICTS 2019*, pp. 103–108, 2019, doi: 10.1109/ICTS.2019.8850984.

[21] X. Chen, R. Al, C. R. Behera, and G. Sin, "Process Synthesis, Design, and Control of Wastewater Treatment Plants," *Ref. Modul. Chem. Mol. Sci. Chem. Eng.*, 2018, doi: 10.1016/B978-0-12-409547-2.14345-8.