

A Comparative Analysis of Rough Sets for Incomplete Information System in Student Dataset

Rd Rohmat Saedudin[#], Shahreen Kasim^{*}, Hairulnizan Mahdin^{*}, Iwan Tri Riyadi Yanto^{##}

[#]*School of Industrial Engineering, Telkom University, 40257 Bandung, West Java, Indonesia*
E-mail: rdrohmat@telkomuniversity.ac.id

^{*}*Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia*
BatuPahat, 86400, Johor, Malaysia
E-mail: {shahreen, hairuln}@uthm.edu.my

^{##}*Department of Information Systems, Universitas Ahmad Dahlan, 55161 Yogyakarta, Indonesia*
E-mail: yanto.itr@is.uad.ac.id

Abstract— Rough set theory is a mathematical model for dealing with the vague, imprecise, and uncertain knowledge that has been successfully used to handle incomplete information system. Since we know that in fact, in the real-world problems, it is regular to find conditions where the user is not able to provide all the necessary preference values. In this paper, we compare the performance accuracy of the extension of rough set theory, i.e. Tolerance Relation, Limited Tolerance Relation, Non-Symmetric Similarity Relation and New Limited Tolerance Relation of Rough Sets for handling incomplete information system in real-world student dataset. Based on the results, it is shown that New Limited Tolerance Relation of Rough Sets has outperformed the previous techniques.

Keywords— rough set theory; incomplete information system; tolerance relation; NLTRS

I. INTRODUCTION

Data mining is the process that include the collection, use historical data to find regularities, patterns, relationships in large data sets and represent the new knowledge to make the data understandable [1]. It has been widely used and has solved many real-life problems [2–4]. In fact, besides enhancing the method used in the learning system [5], data mining has also been used in the field of education to better understand how students learn and identify the settings in which they learn to improve educational outcomes [6].

One of the popular and viable approach for data mining is using rough set theory. Rough set theory [7] was first initiated by Professor Zdzislaw Pawlak as information system i.e. pairs (U, A) where U is the universe of objects, while A is the set of conditional attributes. In recent years, it has been successfully applied to knowledge discovery, information system analysis, artificial intelligence, decision analysis, pattern recognition, etc. Various real-life applications of rough set theory have shown its usefulness

in many domains, i.e. clustering [8–11], modeling conflict analysis [12] with improvement by using soft set theory [13] in terms of computational time, association rule[14], medical diagnosis [15], [16], supplier and distributor selection on Supply Chain Management [17], [18], etc.

The rough set theory does not need any initiatory or supplementary information of data. That is its main advantage. The standard rough set theory, however, can only be implemented to deal with the problems of complete information system where all available objects in the system have attribute values. It is based on the indiscernibility relation that conforms with the reflexive, symmetric and transitive properties. Even though in fact, in the real-world problems, it is regular to find conditions where the user is not able to provide all the necessary preference values, and thus, we need to deal with the incomplete information system.

To cope with this incomplete information system, a lot of effort has been run in studying this domain. The simplest method to deal with incomplete information system is to remove the objects with unknown or missing values [19].

However, this approach certainly reduces the sample size of data. Apart from that, a well-known approach is the extension of rough set theory, called tolerance relation [20]. Yet the disadvantage is, tolerance relation approach leads to poor results in terms of approximation. Subsequently, Stefanowski and Tsoukias [21], [22] proposed similarity relation to improve the results obtained by means of tolerance relation technique. However, Wang [23] and Yang *et al.*, [24] showed that when applying similarity relation, some information are lost, hence they introduced limited tolerance relation. Nevertheless, some information are also lost since limited tolerance relation does not consider the similarity precision between two objects. Nguyen *et al.*, [25] improved the tolerance relation by considering the probability matching between two objects. After all, first, we need to know the probability distribution of the data. Consequently, Deris *et al.* [26] proposed New Limited Tolerance Relation of Rough Sets. This proposal is based on limited tolerance relation by taking into consideration the similarity precision between two objects. The similarity precision is defined when a threshold value is given. Therefore, this study aims to address multiple techniques for incomplete information system based on rough set theory using real-world student data with missing or incomplete values.

This study examines the results of an experimental and comparative analysis on several incomplete information system techniques, including Tolerance Relation (TR), Limited Tolerance Relation (LTR), Non-Symmetric Similarity Relation (NSSR) and New Limited Tolerance Relation of Rough Set (NLTRS) regarding their performance in terms of accuracy. The student data to be analyzed were obtained from the Directorate of Information Systems (SISFO), Telkom University.

The other section of the paper is organized as follows. In Section 2, the basic notion of rough set theory and incomplete information system of the rough set theory are introduced. Afterward, the tolerance relation, limited tolerance relation, non-symmetric similarity relation and new limited tolerance relation of the rough set for handling incomplete information system are briefly described. Section 3 elaborates the results and compares them in terms of accuracy. Finally, the conclusion of this work is presented in Section 4.

II. MATERIALS AND METHODS

In this section, we recalled the notion of information systems, the idea of rough set theory and continued with the essential definitions of incomplete information system techniques based on rough set theory, namely Tolerance Relation, Non-Symmetric Similarity Relation, Limited Tolerance Relation and New Limited Tolerance Relation of Rough Sets.

A. Information System

The idea of information system gives an appropriate tool for the representation of objects in terms of their attribute values. An information system is a 4-tuple (quadruple) $S = (U, A, V, f)$, where $U = \{u_1, u_2, \dots, u_{|U|}\}$ is a non-empty finite set of objects, $A = \{a_1, a_2, \dots, a_{|A|}\}$ is a non-empty finite

set of attributes, $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute a , $f: U \times A \rightarrow V$ is an information function such that $f(u, a) \in V_a$, for every $(u, a) \in U \times A$, called information (knowledge) function [27]. If U in $S = (U, A, V, f)$ contains at least one object with an unknown or missing value, then S is called incomplete information system. The unknown or missing value is denoted by “*” in an incomplete information system. In this paper, we used the quadruple $S^* = (U, A, V, f)$ to denote an incomplete information system. After the idea of an information system was presented as above, we recalled the notion of rough set theory in the following section.

B. Rough Set Theory

The idea of rough set theory was founded on the assumption that every object of the universe of discourse can be associated with some information (data, knowledge). Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The mathematical basis of a rough set theory is that similarity (indiscernible) relationship. The basic set of any set of all similar objects will form the basic grains (atoms) of knowledge of the universe. Some of the basic sets that form a union are called exact - otherwise those are called rough set (not exact, not clear). On a rough set there are objects in boundary line whose certainty cannot be classified, using existing knowledge, as a member of its set or complement.

Foremost, we recalled some fundamental definitions of rough set theory. Formal definitions and detailed description of rough set theory are originated from [7].

The concept of an information table is a quadruple $S = (U, A, V, f)$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, V is the union of attribute domains such that $V = \bigcup_{a \in A} V_a$ for V_a denotes the value domain of attribute a , any $a \in A$ determines a function $f_a: U \rightarrow V_a$ where V_a is the set of values of a .

Two elements $x, y \in U$ in $S = (U, A, V, f)$ are said to be B-indiscernible (indiscernible by the set of the attribute $B \subseteq A$ in S) if and only if $f(x, a) = f(y, a)$, for every $a \in B$ [27]. An indiscernible relation induced by the set of attribute B , denoted by $IND(B)$, is an equivalence relation. It is well known that an equivalence relation can induce a unique partition. The partition of U induced by $IND(B)$ in $S = (U, A, V, f)$ is denoted by U/B and the equivalence class in the partition U/B contains $x \in U$ and denoted by $[x]_B$. Let B be any subset of A in S and let X be any subset of U , the B-lower approximation of X , denoted by $\underline{B}(X)$ and B-upper approximation of X , denoted by $\overline{B}(X)$ respectively, are defined by

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\} \text{ and } \overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\}.$$

The accuracy of approximation of any subset $X \subseteq U$ with respect to $B \subseteq A$, denoted by $\alpha_B(X)$ is measured by $\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}$, where $|X|$ denotes the cardinality of

X. For the empty set ϕ , it is defined as $\alpha_B(\phi)=1$ [28]. Clearly, $0 \leq \alpha_B(X) \leq 1$. If X is a union of some equivalence classes of U, then $\alpha_B(X)=1$. Thus, the set X is exact with respect to B. And, if X is not a union of some equivalence classes of U, then $\alpha_B(X) < 1$. Thus, with respect to B, the set X is not exact. This means that the higher the accuracy of approximation of any subset $X \subseteq U$, the more precise (the less imprecise) it would be [29].

C. Tolerance Relation

Given a complete decision system $S=(U, A, V_*, f)$, where $A=C \setminus \{d\}$, C is a set of condition attributes and d is the decision attribute, such that $f:U \times A \rightarrow V$, for any $a \in A$, where V_a is called domain of attribute a. An incomplete information system $S^*=(U, A, V_*, f)$, for any subset $B \subseteq C$, the tolerance relation T is determined by the following definition.

Definition 1. (See [18,27]) Let $S^*=(U, A, V_*, f)$ be an incomplete information system. A tolerance relation TR is defined as

$$\forall_{x,y \in U} TR(x,y) \Leftrightarrow \forall_{c_j \in B} (c_j(x)=c_j(y) \vee c_j(x)=* \vee c_j(y)=*) \quad (1)$$

Thus,

$$TR = \{(x,y) | x \in U \wedge y \in U \wedge \forall c_j$$

$$(c_j \in B \rightarrow (c_j(x)=c_j(y) \vee c_j(x)=* \vee c_j(y)=*))\}$$

Obviously, TR is reflexive and symmetric but does not need to be transitive. From Definition 1, we described the notion of tolerance class as follows:

Definition 2. Let $S^*=(U, A, V_*, f)$ be an incomplete information system. The tolerance class $I_B^T(x)$ of an object x with reference to an attribute set B is defined as follows:

$$I_B^T(x) = \{y | y \in U \wedge TR_B(x,y)\}. \quad (2)$$

From Definition 2, the notion of lower and upper approximations of tolerance class are described as follows:

Definition 3. Let $S^*=(U, A, V_*, f)$ be an incomplete information system. The lower approximation x_B^T and upper approximation x_T^B of an object set X with reference to attribute set B respectively can be defined as follows:

$$x_B^T = \{x | x \in U \wedge I_B^T(x) \subseteq X\}$$

and

$$x_T^B = \{x | x \in U \wedge I_B^T(x) \cap X \neq \phi\} \quad (3)$$

TABLE I

AN INCOMPLETE INFORMATION TABLE

U / A	a ₁	a ₂	a ₃	a ₄	d
P ₁	3	2	1	0	θ
P ₂	2	3	2	0	θ
P ₃	2	3	2	0	α
P ₄	*	2	*	1	θ
P ₅	*	2	*	1	α
P ₆	2	3	2	1	α
P ₇	3	*	*	3	θ
P ₈	*	0	0	*	α
P ₉	3	2	1	3	α
P ₁₀	1	*	*	*	θ
P ₁₁	*	2	*	*	α
P ₁₂	3	2	1	*	θ

We can represent the above ideas by using an incomplete information system of Wang [23].

Example 1. Table 1 is an incomplete information system, where p_1, p_2, \dots, p_{12} are objects. The a_1, a_2, a_3, a_4 are four condition attributes, where their domain values are $\{0,1,2,3\}$. The d is a decision attribute, where its domain values is $\{\theta, \alpha\}$, $\theta = \{p_1, p_2, p_4, p_7, p_{10}, p_{12}\}$ and $\alpha = \{p_3, p_5, p_6, p_8, p_9, p_{11}\}$.

D. Non-Symmetric Similarity Relation

An object x is considered to be similar to object y only if all their known attribute values are the same. Thus, one object may have more complete description than the other, the inverse relation does not hold [30]. The notion of a non-symmetric similarity relation is given as follows:

Definition 4 (See [30], [31]). Let $S^*=(U, A, V_*, f)$ be an incomplete information system. A non-symmetric similarity relation S is defined as

$$\forall_{x,y \in U} (S_B(x,y) \Leftrightarrow \forall_{c_j \in B} (c_j(x)=c_j(y) \vee c_j(x)=*)) \quad (4)$$

It is clear that S is transitive and reflexive but not symmetric. From Definition 4, we can induce two similarity sets as given in Definitions 5 and 6 below.

Definition 5. Let $S^*=(U, A, V_*, f)$ be an incomplete information system and $B \subseteq A$. The set of objects similar to object x denoted by $Sim_B(x)$ is defined as

$$Sim_B(x) = \{y | y \in U \wedge S_B(y,x)\} \quad (5)$$

Definition 6. Let $S^*=(U, A, V_*, f)$ be an incomplete information system and $B \subseteq A$. The set of objects which x is similar to $Sim_B^{-1}(x)$ is defined as

$$\text{Sim}_B^{-1}(x) = \{y \mid y \in U \wedge S_B(x, y)\} \quad (6)$$

Obviously, $\text{Sim}_B(x)$ and $\text{Sim}_B^{-1}(x)$ are two different sets.

Furthermore, from Definitions 5 and 6, the lower approximation and upper approximation of objects set X can be defined as follows:

Definition 7. Let $S^* = (U, A, V_*, f)$ be an incomplete information system and $B \subseteq A$. The lower-approximation X_B^S and the upper-approximation X_B^B of an object set X with respect to an attribute set $B \subseteq A$ are respectively defined as

$$X_B^S = \{x \mid x \in U \wedge \text{Sim}_B^{-1}(x) \subseteq X\}$$

And

$$X_B^B = \bigcup \{\text{Sim}_B(x) \mid x \in X\} \quad (7)$$

The approximations showed by the non-symmetric similarity relation are more informative than those resulted by tolerance relation.

E. Limited Tolerance Relation

In an information system, two objects may be distinct because of a little missing information. For example, two objects $a = \{x, *, y, z, w\}$ and $b = \{*, v, y, z, w\}$ are similar, but they do not satisfy the non-symmetric similarity relation. To avoid such problem, Wang [22] developed a limited tolerance relation based on the following definition.

Definition 8 (See [23]). Let $S^* = (U, A, V_*, f)$ be an incomplete information system, a subset $B \subseteq A$, and $P_B(x) = \{b \mid b \in B \wedge b(x) \neq *\}$. A binary relation L (limited tolerance relation) defined on U is given by

$$\begin{aligned} \forall_{x, y \in U \times U} (L_B(x, y) \Leftrightarrow \forall_{b \in B} (b(x) = b(y) = *) \vee \\ ((P_B(x) \cap P_B(y) \neq \emptyset) \wedge \\ \forall_{b \in B} ((b(x) \neq *) \wedge (b(y) \neq *) \rightarrow (b(x) = b(y)))))) \quad (8) \end{aligned}$$

Obviously, the limited tolerance relation is symmetric and reflexive but not transitive. In Definition 8, the condition that $(b(x) \neq *) \wedge (b(y) \neq *) \rightarrow (b(x) = b(y))$ is equivalent to $(b(x) = *) \vee (b(y) = *) \vee (b(x) = b(y))$. Thus, two objects that satisfy the tolerance relation but not limited tolerance relation are only those with the formula $P_B(x) \cap P_B(y) = \emptyset$.

In other words, there are two cases where two objects are in a limited tolerance relationship, i.e. if two objects lose all attribute values, and the second case is where there is at least an attribute having an ordinary value for both objects and the two objects have the same value for those attributes. The notion of limited tolerance class is given as follows:

Definition 9. Let $S^* = (U, A, V_*, f)$ be an incomplete information system and a subset $B \subseteq A$. The limited tolerance class is defined as $I_B^L(x) = \{y \mid y \in U \wedge L_B(x, y)\}$.

From Definition 9, the notions of lower approximation and upper approximation of an object x based on the limited tolerance class are given in the following definition.

Definition 10. The lower approximation and the upper approximation of an object x based on the limited tolerance class $I_B^L(x)$ are respectively defined as:

$$D_L^B = \{x \mid x \in U \wedge I_B^L(x) \cap D \neq \emptyset\}$$

and

$$D_B^L = \{x \mid x \in U \wedge I_B^L(x) \subseteq D\}. \quad (9)$$

F. New Limited Tolerance Relation of Rough Sets

An incomplete information system $S^* = (U, A, V_*, f)$, where $A = C \cup \{d\}$, C is a set of condition attributes and d the decision attribute, such that $f: U \times A \rightarrow V_*$. For any $a \in A$, where V_a is called domain of an attribute a and a subset $B \subseteq C$, the *similarity precision* is defined as follows:

Definition 11 (See [26]). Let $P_B(x) = \{b \mid b \in B \wedge b(x) \neq *\}$, the similarity precision δ , is defined as:

$$\delta(x, y) = \frac{|P_B(x) \cap P_B(y)|}{|C|}, \quad (10)$$

where $|\bullet|$ represents the cardinality of the set.

From Definition 11, it is clear that $0 < \alpha(x, y) \leq 1$. Meanwhile, the limited tolerance relation with similarity precision is given as follows:

Definition 12. Given an incomplete information system $S^* = (U, A, V_*, f)$. The limited tolerance relation with similarity precision $L\delta$ is defined as follows:

$$\begin{aligned} \forall_{x, y \in U \times U} (L\delta_B(x, y) \Leftrightarrow \forall_{b \in B} (b(x) = b(y) = *) \vee ((\alpha(x, y)) \geq \delta) \wedge \\ \forall_{b \in B} (((b(x) \neq *) \wedge (b(y) \neq *) \rightarrow (b(x) = b(y)))))) \quad (11) \end{aligned}$$

where $\delta \in (0, 1]$ is a threshold value.

Since $\delta \in (0, 1]$, then $0 < \alpha(x, y) \leq 1$ which implies that $P_B(x) \cap P_B(y) \neq \emptyset$ holds, but not vice versa if the particular threshold value of the similarity is given.

Afterward, we recalled the extended tolerance relation by using similarity precision with a threshold.

Definition 13. Given an incomplete information system $S^*=(U, A, V_s, f)$, a subset $B \subseteq C$ and a threshold δ . The limited tolerance relation with similarity precision is defined as follows:

$$L\delta_B(x, y) \Leftrightarrow \alpha_B(x, y) \geq \delta \quad (12)$$

In the next section, we discussed the performance of four (4) techniques in terms of accuracy.

III. RESULT AND DISCUSSION

In this section, we compare all the incomplete information system techniques based on accuracy. A real-world dataset that contains incomplete missing values is used. This dataset was obtained from the Directorate of Information Systems (SISFO), Telkom University. It contains 200 instances and eight (8) categorical attributes. The attributes that have been used are Student ID, 1st GPA, 2nd GPA, 3rd GPA, 4th GPA, 5th GPA, 6th GPA, and Performance of Student. Here, irrelevant attributes such as name, gender, student residential address, etc. have been removed. The occurrence of missing values might be due to several possibilities, such as the student was on leave, the GPA score is not final, the student is not enrolled in certain semester, etc.

The Performance Status field represents the performance of students during their studies. The description of each attribute of the dataset is shown in Table 2 as follows:

TABLE II
DESCRIPTION OF DATASET ATTRIBUTES

Attribute Name	Description	Attribute Set Value
ID	ID of student	{1,2,3,...,200}
1st	Letter representation of GPA of student in first semester	{A, AB, B, BC, C, D, E}
2nd	Letter representation of GPA of student in second semester	{A, AB, B, BC, C, D, E}
3rd	Letter representation of GPA of student in third semester	{A, AB, B, BC, C, D, E}
4th	Letter representation of GPA of student in fourth semester	{A, AB, B, BC, C, D, E}
5th	Letter representation of GPA of student in fifth semester	{A, AB, B, BC, C, D, E}
6th	Letter representation of GPA of student in sixth semester	{A, AB, B, BC, C, D, E}
Performance	Performance status of student	{POSITIVE, NEGATIVE}

The values of GPA are in the form of letter representation of their actual numeric score (4.0 scale). The conversion of the actual score of GPA to a letter representation based on a

standard that is implemented by Telkom University is as shown in Table 3 below:

TABLE III
CONVERSION OF GPA

Range of GPA	GPA Letter	Category
3.51 – 4.0	A	Excellent
3.01 -3.5	AB	Very Good
2.51-3.0	B	Good
2.01-2.5	BC	Fair
1.51-2.0	C	Satisfactory
1.1 - 1.5	D	Passing
0.0 – 1.0	E	Poor

The sample of 10 out of 200 of student data that are used as a dataset in this paper is shown in Table 4 as follows:

TABLE IV
SAMPLE OF DATASET

ID	1st	2nd	3rd	4th	5th	6th	Performance
1	B	B	AB	B	*	*	POSITIVE
2	B	BC	BC	AB	AB	A	POSITIVE
3	B	BC	BC	*	BC	B	NEGATIVE
4	B	C	C	D	B	*	NEGATIVE
5	D	B	BC	C	*	*	NEGATIVE
6	AB	B	AB	A	AB	A	POSITIVE
7	D	C	C	BC	D	BC	NEGATIVE
8	BC	B	B	B	*	*	NEGATIVE
9	AB	B	AB	AB	AB	A	POSITIVE
10	A	A	AB	B	*	*	POSITIVE

In order to apply all the techniques, the experiments are developed using MATLAB version 7.14.0.334 (R2012a). The techniques are executed sequentially on a processor Intel 1.5 GHz CPUs, with total main memory 2G of RAM and the operating system is Windows 7. The computation results comparing all four (4) techniques in terms of accuracy are shown in Figure 1.

Based on the above result, we can see that all techniques have recorded low accuracy, but it is undeniable that New Limited Tolerance Relation of Rough Sets has outperformed other techniques, i.e. Tolerance Relation, Limited Tolerance Relation, and Non-Symmetric Similarity Relation techniques. We have observed that it has low accuracy due to the properties of the dataset, e.g. the data are homogeneous, and because of the small number of instances used, which are only 200 instances. Although all of the techniques have achieved low accuracy, the improvement of NLTRS is very significant. The percentage of improvement of NLTRS as compared to TR, LTR and NSSR in terms of accuracy is shown in Table 5 below.

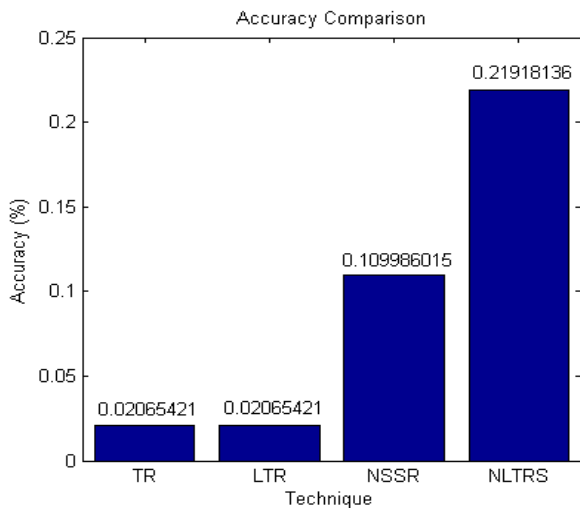


Fig. 1. Accuracy Comparison of each technique

TABLE V
PERCENTAGE OF IMPROVEMENT

No.	Technique	Result	Improvement % of NLTR to others
1	TR	0.02065421	90.57 %
2	LTR	0.02065421	90.57 %
3	NSSR	0.109986015	49.82 %
4	NLTRS	0.21918136	-
Improvement Average			76.99%

The improvement performance of NLTRS compared to other techniques is calculated by the following formula:

$$Impr (\%) = \frac{(NLTRS - other\ techniques)}{NLTRS} \times 100\%$$

(13)

In summary, based on experiments that we have carried out using the real-world student dataset, the NLTRS achieved higher accuracy than the other techniques. The improvement average of NLTRS as compared other techniques in terms of accuracy of NTRS is 76.99%.

IV. CONCLUSION

A number of techniques of extended rough set theory for handling incomplete information system have been proposed i.e. Tolerance Relation, Limited Tolerance Relation, Non-Symmetric Similarity Relation and New Limited Tolerance Relation of Rough Sets. However, all of the techniques have not been implemented in the real-world dataset that contains missing values or incomplete information system. In this paper, the researchers have applied all of the techniques and compared the performance of each technique in terms of accuracy. From the results, it is illustrated that New Limited Tolerance Relation of Rough Set achieved higher accuracy as compared to Tolerance Relation, Limited Tolerance Relation, and Non-Symmetric Similarity Relation with the improvement average of 76.99%.

REFERENCES

- [1] U. M. Fayyad, "Data mining and knowledge discovery: Making sense out of data," IEEE Expert: Intelligent Systems and Their Applications, vol. 11, no. 5, pp. 20–25, 1996.
- [2] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students," in Educational Data Mining 2008, 2008.
- [3] S. Zainudin, D. S. Jasim, and A. A. Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, no. 6, pp. 1148–1153, 2016.
- [4] M. K. Nasution, "Social Network Mining (SNM): A Definition of Relation between the Resources and SNA," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, no. 6, 2016.
- [5] M. S. Khalid, E. Sutoyo, E. N. Sari, T. Herawan, and others, "Teaching and Learning Using Computers: How Should We Tread on Its' Changing Technology?," International Journal of Emerging Technologies in Learning, vol. 9, no. 5, 2014.
- [6] C. Romero and S. Ventura, "Data mining in education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12–27, 2013.
- [7] Z. Pawlak, "Rough sets," International Journal of Computer & Information Sciences, vol. 11, no. 5, pp. 341–356, 1982.
- [8] I. T. R. Yanto, P. Vitasari, T. Herawan, and M. M. Deris, "Applying variable precision rough set model for clustering student suffering study's anxiety," Expert Systems with Applications, vol. 39, no. 1, pp. 452–459, 2012.
- [9] R. R. Saedudin, S. B. Kasim, H. Mahdin, and M. A. Hasibuan, "Soft Set Approach for Clustering Graduated Dataset," in International Conference on Soft Computing and Data Mining, 2016, pp. 631–637.
- [10] D. W. Jacob, M. F. M. Fudzee, M. A. Salamat, R. R. Saedudin, I. T. R. Yanto, and T. Herawan, "An Application of Rough Set Theory for Clustering Performance Expectancy of Indonesian e-Government Dataset," in International Conference on Soft Computing and Data Mining, 2016, pp. 638–646.
- [11] I. T. R. Yanto, R. R. Saedudin, D. Hartama, and T. Herawan, "Clustering Based on Classification Quality (CCQ)," in International Conference on Soft Computing and Data Mining, 2016, pp. 327–335.
- [12] Z. Pawlak, "Some remarks on conflict analysis," European Journal of Operational Research, vol. 166, no. 3, pp. 649–654, 2005.
- [13] E. Sutoyo, M. Mungad, S. Hamid, and T. Herawan, "An Efficient Soft Set-Based Approach for Conflict Analysis," PloS one, vol. 11, no. 2, p. e0148837, 2016.
- [14] D. W. Jacob, M. F. M. Fudzee, M. A. Salamat, R. Saedudin, Z. Abdullah, and T. Herawan, "Mining Significant Association Rules from on Information and System Quality of Indonesian E-Government Dataset," in International Conference on Soft Computing and Data Mining, 2016, pp. 608–618.
- [15] S. Tsumoto, "Modelling medical diagnostic rules based on rough sets," in International Conference on Rough Sets and Current Trends in Computing, 1998, pp. 475–482.
- [16] B. Tripathy, "Application of rough set based models in medical diagnosis," in Handbook of Research on Computational Intelligence Applications in Bioinformatics, IGI Global, 2016, pp. 144–168.
- [17] Z. Zou, T.-L. B. Tseng, H. Sohn, G. Song, and R. Gutierrez, "A rough set based approach to distributor selection in supply chain management," Expert Systems with Applications, vol. 38, no. 1, pp. 106–115, 2011.
- [18] J. Chai and J. N. Liu, "A novel believable rough set approach for supplier selection," Expert systems with applications, vol. 41, no. 1, pp. 92–104, 2014.
- [19] B. P. Bunting, G. Adamson, and P. K. Mulhall, "A Monte Carlo examination of an MTMM model with planned incomplete data structures," Structural Equation Modeling, vol. 9, no. 3, pp. 369–389, 2002.
- [20] M. Kryszkiewicz, "Rough set approach to incomplete information systems," Information sciences, vol. 112, no. 1, pp. 39–49, 1998.
- [21] J. Stefanowski and A. Tsoukias, "Incomplete information tables and rough classification," Computational Intelligence, vol. 17, no. 3, pp. 545–566, 2001.
- [22] J. Stefanowski and A. Tsoukiàs, "On the extension of rough sets under incomplete information," in International Workshop on Rough

- Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, 1999, pp. 73–81.
- [23] G. Wang, “Extension of rough set under incomplete information systems,” in *Fuzzy Systems, 2002. FUZZ-IEEE’02. Proceedings of the 2002 IEEE International Conference on*, 2002, vol. 2, pp. 1098–1103.
- [24] X. Yang, “An improved model of rough sets on incomplete information systems,” in *Management of e-Commerce and e-Government, 2009. ICMECG’09. International Conference on*, 2009, pp. 193–196.
- [25] D. Van Nguyen, K. Yamada, and M. Uehara, “Extended tolerance relation to define a new rough set model in incomplete information systems,” *Advances in Fuzzy Systems*, vol. 2013, p. 9, 2013.
- [26] M. M. Deris, Z. Abdullah, R. Mamat, and Y. Yuan, “A new limited tolerance relation for attribute selection in incomplete information systems,” in *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*, 2015, pp. 964–970.
- [27] J. Dai, W. Wang, Q. Xu, and H. Tian, “Uncertainty measurement for interval-valued decision systems based on extended conditional entropy,” *Knowledge-Based Systems*, vol. 27, pp. 443–450, 2012.
- [28] D. Parmar, T. Wu, and J. Blackhurst, “MMR: an algorithm for clustering categorical data using rough set theory,” *Data & Knowledge Engineering*, vol. 63, no. 3, pp. 879–893, 2007.
- [29] T. Herawan, M. M. Deris, and J. H. Abawajy, “A rough set approach for selecting clustering attribute,” *Knowledge-Based Systems*, vol. 23, no. 3, pp. 220–231, 2010.
- [30] J. Zhou and X. Yang, “Rough set model based on hybrid tolerance relation,” in *International Conference on Rough Sets and Knowledge Technology*, 2012, pp. 28–33.
- [31] Q. Zhou, “Research on Tolerance-Based Rough Set Models,” in *System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 International Conference on*, 2010, vol. 2, pp. 137–139.