ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING FUZZY SOFT SET ASSOCIATION RULE MINING APPROACH TO IDENTIFY THE STUDENT SKILL DATA ASSOCIATION

[1]DEDE ROHIDIN, [2]NOOR A. SAMSUDIN, [3]MOHD F. AB. AZIZ, [4]SHAMSUL K. A. KHALID

[1]Telkom University, Department of Computer Science, Indonesia

[2,3,4]Universiti Tun Hussein Onn. Computer Science and Information Technology, Malaysia

E-mail: [1]dede.rohidin@tass.telkomuniversity.ac.id, [2] azah@uthm.edu.my

ABSTRACT

Graduates shall see the importance of improving their skills to prepare themselves to get the right jobs. Considering such important requirements, therefore many universities have designed various skill development programs in their curriculum. The students enrol in these programs and eventually, their skills performance will be evaluated using certain scores. This research does not aim to calculate the scores. Instead, we focus on how to find the relationship between parameters presented for the evaluation. In this study, we use the fuzzy-soft-association-rule mining (FSAR) approach and proposed the fast algorithm for finding association rule on fuzzy soft set. FSAR is a tool that combines Fuzzy Soft Set concepts and Association Rule Mining. We found that FSAR is an effective method to describe the relationship between parameters in large size data. Using FSAR, we will find a significant parameter or uncorrelated parameters for further analysis. This study recommends the selected parameters for determining the type of training for students. By selecting the right training, the University can reduce training cost significantly.

Keywords: *Association Rules, Soft Set, Fuzzy-Soft-Association Rule*.

1. INTRODUCTION

Association rules and soft set method are two famous methods in knowledge data discovery (KDD) that are often widely used. Association rule proposed by Agrawal[1]. At the first time, association rules used for market basket analysis. Current, association rules have been widely used in various disciplines. One of the significant association rule mining applications is about determining data correlation in a large data set. There are association rule applications in data categorization or data classification problems [2][3][4][5]. Martin [6] proposed a business intelligence model based on association rule mining algorithm and financial domain ontology to predict bankruptcy. Reagan Hu [7] studied the Apriori algorithm of Association Rules and its deficiencies to implement breast disease diagnosis. Zhi-Min Xu [8]applied the concept of association rules for revenue analysis from financial transactions.

The soft set [9] is known for its reliability in handling crisp parameter but still has challenges in dealing with real numbers. Therefore, Maji [10] proposed a fuzzy soft set for handling real number parameters. As an association rule, the fuzzy soft set has been implemented in various disciplines. Handaga [11] used a fuzzy soft set theory to develop numerical data classification algorithm. Mangalampalli [12] implemented fuzzy association rule mining on large data set to improve speed performance. Jiang [13] used ontology based on soft set for semantic decision making.

There are also association rule applications in soft set data representation reported in the literature. Dede rohidin [14] proposed fuzzy soft set association rules, a method that combines fuzzy soft set and association rules (FSAR). FSAR enables data correlation finding in a large data set.

In this study, we will use FSAR in the psychology area. We extract the student soft skill data, the soft-kill needed by graduated. Coz, many employers require not only high qualifications but also a graduate must possess appropriate skills and competencies[15]. We know two types of skills: soft skill and hard skills. Snyder [16] define hard skill as a technical skill. Rainsbury [17] define soft skills as behaviour required to apply technical skills.

Certainly, we need both, hard skill and soft skill to perform In fact, some job positions prefer the soft skills required to be more fulfilled rather than the hard skill requirement [18].

There are some soft skills needed by a graduate: active learning, social perceptiveness, teamwork, etc. In this research, we find out the relation between different kinds of skills required by graduates. There are eighteen kinds of skills to be studied. The value of each skill is achieved by assessment of 40 students. The assessment is a performance assessment that is guided by several experts. We believe that of the eighteen types of skills tested are mutually influential skills. For that reason, in this study, we sought the relationship between the types of Skill. By looking at the relationship between the skills, we can choose a certain significant skill. Significant skills can be used as a basis to determine as little as possible the type of training that will be given to students so they can have as many skills as possible. The process of finding the relation between these skills is done using the Fuzzy soft set Association rule (FSAR) method. This method uses the Support and confidence parameters to measure the relationship between the Skill.

This paper has extended the approach presented in [14] with the following contributions:

a. Extension of fuzzy soft set association rules concept in psychology area.
b. Implementation of a fuzzy soft association rules concept through some derivation steps.
c. Propose a fast algorithm for finding association rules on fuzzy soft set.

The discussion in this paper starts from the introduction presented in section 1. Next on section 2 presented the concept of association rule as well as a soft set that is the basis of the different methods will be introduced. Section 3 presents the proposed method: Fuzzy Soft Association Rules Concept and the algorithm for finding rules. Section 4 discusses the implementation of fuzzy soft association rules on student data set. Finally, a conclusion and future work are presented at the end of the discussion.

2. LITERATURE REVIEW

In this section, we describe the concept and notation association rule of Agrawal [1], soft set of Molodtsov [9] and fuzzy soft set of R. K. Maji [10]. Those concepts used in our research.

2.1 Associations Rules

Association rule is a tool to discover the relationships between certain data in the itemset. Analysis of the association rule is often also called market basket Analysis. Formally, Association rule mining defined as follows: $I = \{i_1, i_2,\ldots, i_m\}$ is attribute set and i_j where j = 1, 2, ..., m, is Attribute item. Let, set of transaction D = $\{ t_1, t_2,\ldots, t_m\}$, where each transaction t has a t-itemset, that is $t_k = \{t_{k1}, t_{k2},\ldots, t_{kn}\} \subseteq I$. An association rule between items set X and Y, denoted by the form $X \Rightarrow Y$, where $X \subseteq$ I, $Y \subseteq$I and $X \cap Y = \boldsymbol{\phi}$. X is called antecedent and Y is called *consequent*.

In the association rule, there is some measure of confidence in determining the level of association between itemset obtained from processing results in a certain way. The size of the trust is Support and confidence.

Support shows how the dominance of an itemset to the overall transaction. Support an itemset X revealed with sup(x) is the number of transactions containing X divided by the total number of transactions in D. While support for association rule X⇒Y is defined as the number of transactions containing X and Y divided by all of the transactions.

$$sup(X\Rightarrow Y)= \frac{Transaction\ Containing\ both\ X\ and\ Y}{Total\ Number\ of\ Transaction}$$

The Confidence is a quantity that indicates the magnitude of the relationship between the 2 itemsets conditionally. Confidence X⇒Y stated *Conf*(X⇒Y) is a measure that indicates how often the items Y appears in the transaction consisting of the X. Confidence X⇒Y is the ratio between the transactions containing both X and Y with transactions that only contains X.

$$conf(X\Rightarrow Y) = \frac{Transaction\ Containing\ both\ X\ and\ Y}{Transaction\ Containing\ X}$$

Example 1:

To illustrate the concept of association rules, the following simple example is given. Suppose there are five purchase transaction data in a supermarket, the data can be seen in table 1. The items set in this case are {coffee, milk, sugar, candy, soap}. In each data transaction, if a_{ij} is a value in a table than a_{ij}=1 means there is a purchase of item j in the transaction

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

i and a_{ij}=0 means there is no purchase of item j in the transaction i.

Table 1: The Simple Data Transaction

Trans	Coffee	milk	sugar	Candy	Soap
1	1	0	1	0	0
2	0	1	0	0	1
3	1	1	1	0	0
4	0	0	0	1	1
5	0	0	1	1	0

An example rule for the supermarket could be {coffee, milk} ⇒ {sugar}. Meaning that, if coffee and milk are bought, customers also buy sugar with

$$sup(\{\text{coffee, milk}\} \Rightarrow \{\text{sugar}\}) = 1/5 = 0.2 = 20\%$$

$$Conf(\{\text{coffee, milk}\} \Rightarrow \{\text{sugar}\}) = 1/1 = 1 = 100\%$$

The example is very simple. In fact, Association rules are mostly implemented in large data and attributes, so special algorithms are needed to speed up the process of finding association rules.

A priori Algorithm

A priori Algorithm proposed by Agrawal is one of a famous algorithm for finding height frequent itemset on a large dataset [19]. Algorithm of A priori showed in figure 2.1:

Fast Algorithm for finding Association Rule

Input: item set
Output: rules of each class

```
L1 ← {large 1 – itemsets}
k ← 2
While Lk-1 ≠ 0
    Ck ← {a ∪ {b} | a ∈ Lk-1 ∧ b ∉ a} –
        {c | {s | s ⊆ c ∧ |s| = k – 1} ⊈ Lk-1}
    for transaction t ∈ T
        Ct ← { c | c ∈ Ck ∧ c ⊆ t }
    for candidates c ∈ Ct
        count [c ] ← count [c ] + 1
Lk ← {c ∈ Ck ∧ count [c] ≥ ε}
k ← k + 1
return ∪k Lk
```

Figure Error! No text of specified style in document.**-1 A priori Algorithm**

This Algorithm is divided into two stages. In the first stage, repeated processes are carried out with the aim of forming itemset that often appears. Scan all parameters as 1-itemset and calculate support for each candidate. Next is the iteration process. In iteration k, form k-itemset by combining the k-1 itemset obtained from the previous process. Select itemset that has support higher than minimum support. The process is repeated until larger itemset cannot be formed again.

In the second stage, calculate the confidence value for all frequent itemset. Next, select the itemset that has a confidence value greater than the minimum confidence.

2.2 Soft Set Theory

Definition 1. : *Let U be an initial universe set and E be a set of parameters. Let P(U) denote the power set of U. a pair (F, E) is called a soft set over U, if and only if F is a mapping of E into P(U),* given by f_E: E →*P*(U).

Thus, a soft set over U can be represented by the collection of ordered pairs

$$F_E = \{(x, f_E(x)) \mid x \in E, f_E(x) \in P(U)\}$$

A soft set (F, E) over the universe *U* can be regarded as a parameterized family of subsets of the universe *U*, which gives an approximate (soft) description of the objects in *U*.

Example 2

Suppose *U* = $\{u_1, u_2, u_3, u_4, u_5, u_6\}$ is the set of houses under consideration, *E* = $\{x_1, x_2, x_3, x_4, x_5\}$ is the set of parameters where x_1 stand for the parameter expensive, x_2 for beautiful, x_3 for wooden, x_4 for cheap and x_5 in the green surroundings. Suppose that observation were obtained : expensive houses = $\{u_2, u_4\}$, beautiful houses = $\{u_1, u_3\}$, wooden houses = $\{u_3, u_4, u_5\}$, cheap houses = $\{u_1, u_3, u_5\}$, in the green

Journal of Theoretical and Applied Information Technology
15th July 2019. Vol.97. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

surroundings = { u_1 }. The attractiveness of the houses which Mr. X is going to buy can describe with soft set F_E where:

$$f_E(x_1) = \{ u_2 , u_4 \}$$
$$f_E(x_2) = \{ u_1 , u_3 \}$$
$$f_E(x_3) = \{ u_3 , u_4 , u_5 \}$$
$$f_E(x_4) = \{ u_1 , u_3 , u_5 \}$$
$$f_E(x_5) = \{ u_1 \}$$

the soft set F_E is

$$F_E = \left\{ \begin{matrix} (x_1, \{ u_2 , u_4 \}), (x_2 , \{ u_1 , u_3 \}), \\ (x_3, \{ u_3 , u_4 , u_5 \}), (x_4, \{ u_1 , u_3 , u_5 \}), (x_5, \{ u_1 \}) \end{matrix} \right\}$$

2.3 Fuzzy Soft Set Theory

The fuzzy soft set is given in the following definition.

Definition 3. Let U be a universe. a fuzzy set X over U is a set defined by a function μ_x representing the mapping. $\mu_x : U \rightarrow [0,1]$

From Definition 3, the mapping μ_x is called the membership function of *X*, and the value $\mu_x(u)$ is called the grade of membership $u \in U$. The value represents the degree of *u* belongs to the fuzzy set *X*. Thus, a fuzzy set *X* over U can be represented as follows:

$$X = \{(\mu_x(u)/u) : u \in U, \mu_x(x) \in [0,1]\}$$

Note that the set of all the fuzzy sets over U will be denoted as $F(U)$. A fuzzy soft set (*fuzzy soft set*) F_A over U is a set defined by a function f_A representing a mapping fA: $A \rightarrow F(U)$ and $f_A(x)$=0, if $x \notin A$, $A \subseteq E$. Thus, the *fuzzy soft set* F_A over U can be represented by the set of ordered pairs

$$F_A = \{(x, f_A(x)) / \ x \in A, f_A(x) \in F(U)\}$$

Note that the set of all fuzzy soft set over *U* will be denoted by $FS(U)$. Let us consider an example of a fuzzy soft as follow:

Example 3:

From example 2, suppose that for detailed observation, we give a weight to the assessment. The "expensive" parameter gives a weight value of 0.5, 1, 0.4, 1, 0.3, 0 for house u respectively then $f_A(x_1)$ is:

$$f_A(x_1) = \{(u_1,0.5), (u_2,1), (u_3,0.4), (u_4,1), (u_5,0.3), (u_6,0)\}$$

the other parameters given as follows:

$$f_A(x_2) = \{(u_1,1), (u_2,0.4), (u_3,1), (u_4,0.4), (u_5,0.6), (u_6,0.8)\}$$

$$f_A(x_3) = \{(u_1,0.2), (u_2,0.3), (u_3,1), (u_4,1), (u_5,1), (u_6,0)\}$$

$$f_A(x_4) = \{(u_1,1), (u_2,0), (u_3,1), (u_4,0.2), (u_5,1), (u_6,0.2)\}$$

$$f_A(x_5) = \{(u_1,0.8), (u_2,0.1), (u_3,0.5), (u_4,0.3), (u_5,0.2), (u_6,0.3)\}$$

The "attractive-ness of the houses" which Mr. X is going to buy can describe by F_A:

$$F_A = \{ (x_1 , \{(u_1,0.5), (u_2,1), (u_3,0.4), (u_4,1), (u_5,0.3), (u_6,0)\}), (x_2 , \{(u_1, 1), (u_2,0.4), (u_3,1), (u_4,0.4), (u_5,0.6), (u_6,0.8)\}), (x_3, \{(u_1,0.2), (u_2,0.3), (u_3,1), (u_4,1), (u_5,1), (u_6,0)\}), (x_4 , \{(u_1,1), (u_2,0), (u_3,1),(u_4,0.2), (u_5,1), (u_6,0.2)\}), (x_5 , \{(u_1,0.8), (u_2,0.1), (u_3,0.5), (u_4,0.3), (u_5,0.2), (u_6,0.3))\}$$

2.4 Matrix Representation for Soft Set

In matrix representation, the F_A can be presented by Table 2 as follow:

Table 2: representation of soft set in table

U	x_1	x_2	x_3	$\cdots$	x_n
u_1	$\mu_{f_{A(x_1)}}(u_1)$	$\mu_{f_{A(x_2)}}(u_1)$	$\mu_{f_{A(x_3)}}(u_1)$	$\cdots$	$\mu_{f_{A(x_n)}}(u_1)$
u_2	$\mu_{f_{A(x_1)}}(u_2)$	$\mu_{f_{A(x_2)}}(u_2)$	$\mu_{f_{A(x_3)}}(u_2)$	$\cdots$	$\mu_{f_{A(x_n)}}(u_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_m	$\mu_{f_{A(x_1)}}(u_m)$	$\mu_{f_{A(x_2)}}(u_m)$	$\mu_{f_{A(x_3)}}(u_m)$	$\cdots$	$\mu_{f_{A(x_n)}}(u_m)$

where $\mu_{f_{A(x_i)}}$ is the membership function of f_A

If $a_{ii} = \mu_{f_{A(x_i)}}(u_j)$ for i =1,2,3...n and j = 1,2,3...m then the f_A is uniquely characterized by a matrix,

Journal of Theoretical and Applied Information Technology
15th July 2019. Vol.97. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$[a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The above matrix which has a size of m rows and n columns is called $m \times n$ of the f_A over U. The matrix of the f_A from Example 2.1 is,

$$F_A = \begin{vmatrix} 0,5 & 1 & 0,2 & 1 & 0,8 \\ 1 & 0,4 & 0,3 & 0 & 0,1 \\ 0,4 & 1 & 1 & 1 & 0,5 \\ 1 & 0,4 & 1 & 0,2 & 0,3 \\ 0,3 & 0,6 & 1 & 1 & 0,2 \\ 0 & 0,8 & 0 & 0,2 & 0,3 \end{vmatrix}$$

2.5 Cardinality of Soft Set

Theorem 1 The cardinal set of FA, denoted by $cF_A = \{\mu_{cF_A}(x)\ /x{:}x \in \mathrm{E}\}$ is a fuzzy soft set over E.

Proof.

Let $F_A \in FS(U)$ with cardinal set $cF_A = \{\mu_{cF_A}(x)/x{:}x \in \mathrm{E}\}$, where μ_{cF_A} is a membership function of cF_A and

$$\mu_{cF_A}(x) = \frac{|f_A(x)|}{|U|} \in [0,1]$$

Where, $|U|$ is the cardinality of the universe U and $|f_A(x)|$ is the scalar cardinality of fuzzy set $f_A(x)$. It is mean $\mu_{cF_A}: E \rightarrow [0,1]$.

The following definition describes the representation of cF_A.

Definition 4 Let $F_A \in FS(U)$ and $cF_A \in cFS(U)$. Assume that $E = \{x_1, x_2, \cdots, x_n\}$ and $A \subseteq E$. The cF_A can be presented in Table 3.5 as follow

Table 3: Representation of cF_A

E	x_1	x_2	…	x_n
$\mu_{cF_A}(x)$	$\mu_{cF_A}(x_1)$	$\mu_{cF_A}(x_2)$		$\mu_{cF_A}(x_n)$

From Definition 4, if $a_{ij} = \mu_{cF_A}(x_j)$ for $j = 1,2,\cdots,n$, then the cardinal set $c\Gamma_A$ is uniquely characterized by a matrix

$$[a_{ij}] = [a_{11} \quad a_{12} \quad \cdots \quad a_{1n}]$$

3. PROPOSED METHODE

3.1 Fuzzy Soft Set Association Rules Theory

A fuzzy soft set association rule (FSAR) is an association rule form of fuzzy soft set representation [14]. Let F_A *and* F_B are fuzzy soft set over U where $A, B \subseteq E$, and $A \cap B = \phi$. Association Rule between F_A and F_B is denoted by the form $F_A \Rightarrow F_B$, *a* fuzzy soft set F_A is called antecedent and F_B *is* consequent. Support and confidence are defined as follow:

$$sup(F_A \Rightarrow F_B) = \frac{|f_A(x) \cap f_B(y)|}{|U|} = \frac{|min(f_A(x), f_B(y))|}{|U|} \quad (1)$$

$$conf(F_A \Rightarrow F_B) = \frac{|f_A(x) \cap f_B(y)|}{|f_A(x)|} = \frac{|\min(f_A(x), f_B(y))|}{|f_A(x)|} = \frac{sup(F_A \Rightarrow F_B)}{sup(F_A)} \quad (2)$$

where $|U|$ is the cardinality of the universe U, and $|f_A(x)|$ is the scalar cardinality of fuzzy set $f_A(x)$

Example 4.

let A = { x_2}, B = {x_1}, where A, B ⊆ E. F_A and F_B *are* two *fuzzy soft set* over U.

A fuzzy soft association rule: $F_A \Rightarrow F_B$

{beautiful (x_2)} ⇒ {expensive (x_1)}

$$Sup(F_A \Rightarrow F_B) = \frac{|0.5,\ 0.4,\ 0.4,\ 0.4,\ 0.3,\ 0|}{|6|} = \frac{0.5 + 0.4 + 0.4 + 0.4 + 0.3 + 0}{6} = 0.33$$

$$Conf(F_A \Rightarrow F_B) = \frac{|0.5,\ 0.4,\ 0.4,\ 0.4,\ 0.3,\ 0|}{|1,\ 0.4,\ 1,\ 0.4,\ 0.6,\ 0.8\,|} = \frac{0.5+0.4+0.4+0.4+0.3+0}{|1+0.4+1+0.4+0.6+0.8\,|}$$

$= \frac{2}{4.2}$

$= 0.48$

The rule means"*beautiful and expensive house*" has supported by 33% of data and a 48% confidence level.

3.2 FSAR Algorithm

FSAR is a new model which effective to describe the relations between parameter in large data. One of the problems in the implementation of the FSAR for large data is the length of processing time. To address this, in this section we will discuss the construction of algorithms to speed up the processing time. This algorithm is a modification of the apriori algorithm.

Given a simple fuzzy soft set data. The data consist of 4 data (u_1, u_2, u_3) with 7 parameters $E=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$. A function f_A representing a mapping f_A:A→FS(U) define as follow:

$f_A(x_1)$ = {(u_1,0.3),(u_2, 0.93),(u_3,0.9),(u_4,0.3)}

$f_A(x_2)$ = {(u_1, 0.54),(u_2,0.54),(u_3,0.93),(u_4,0.78)}

$f_A(x_3)$ = {(u_1,0.7),(u_2,0.5),(u_3,0.68),(u_4,0.78)}

$f_A(x_4)$ = {(u_1,0.2),(u_2,0.9),(u_3,0.9),(u_4,0.45)}

$f_A(x_5)$ = {(u_1,0.5),(u_2,0.75),(u_3,0.48),(u_4,0.73)}

$f_A(x_6)$ = {(u_1,0.4),(u_2,0.4),(u_3,0.431),(u_4,0.5)}

$f_A(x_7)$ = {(u_1,0.2),(u_2,0.2),(u_3,0.2),(u_4,0.4)}

The data can show in form of table. See table 4.

Table 4: The Simple Data

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7
U_1	0.3	0.54	0.7	0.2	0.5	0.4	0.2
U_2	0.93	0.54	0.5	0.9	0.75	0.4	0.2
U_3	0.9	0.93	0.68	0.9	0.48	0.43	0.2
U_4	0.3	0.78	0.78	0.45	0.73	0.5	0.4

Now, we will construct an algorithm to find a relationship between the parameters. The Algorithm divide into 3 stages.

Stage 1. At this step, all parameter as a candidate of 1-itemset (C1). Scan all the parameters and calculate the support value of each parameter, using equations 1. All item set member of the 1-items set is listed in table 5.

Table 5: Candidate 1-item set (C-1)

Items Set	Support
{1}	0.6075
{2}	0.6975
{3}	0.665
{4}	0.6125
{5}	0.615
{6}	0.4325
{7}	0.2500

Defined the threshold of support value (*minsup*) then keep a parameter from C1 if the support value upper than threshold. The item that have support upper than *threshold* called frequent. Remove all item which not frequent. In this simulation, we set *minsup* = 0.5. So, we selected item set {1}, {2}, {3}, {4} and {5} as 1-items set (L1) whereas {6} and {7} removed. All item set L1 showed in table 6.

Table 6: 1-item set (L-1)

Items Set	Support
{1}	0.7725
{2}	0.6975
{3}	0.665
{4}	0.7075
{5}	0.615

Stage 2. The process for this stage is done repeatedly. Started with generating a list of all pairs of frequent items. Example, the Element {1} combined with element {2} be {1 2}. Next, calculated support for {1 2} using equation 1.

A = { x_1}, B = {x_2},

$$sup(F_A \Rightarrow F_B) = \frac{|min(f_A(x), f_B(y))|}{|U|}$$

$$= \frac{|0.3,\ 0.54,\ 0.9, 0.3|}{|4|}$$

$$= \frac{0.3 +\ 0.54 +\ 0.9 +\ 0.3}{4}$$

Journal of Theoretical and Applied Information Technology
15th July 2019. Vol.97. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

= 0.51

Then calculate the support of another element, too. This combined form element candidate 2-itemset (C2). List of C2 can be seen in table .7.

Table 7: Candidate 2-item set (C-2)

Items Set	Support
{1 2}	0.51
{1 3}	0.445
{1 4}	0.575
{1 5}	0.4575
{2 3}	0.625
{2 4}	0.5225
{2 5}	0.5625
{3 4}	0.4575
{3 5}	0.5525
{4 5}	0.47

From candidate 2-items set C2 select frequent item to build 2-items set L2. From C2, support {1 3} = 0.445, support {1 5} = 0.4575 and support {4 5} = 0.47. Those items are not frequent because the support less than minsupp = 0.5. The items set that is not frequent must be removed. So, 2-items set L2 are {1 2}, {1 4}, {2 3}, {2 4}, {2 5} and {3 5}. List of 2-items Set L2 can be seen in a table 8.

Table 8: 2-item set (L-2)

Items Set	Support
{1 2}	0.51
{1 4}	0.575
{2 3}	0.625
{2 4}	0.5225
{2 5}	0.5625
{3 5}	0.5525

Furthermore, the above process is repeated to build candidate items Set C3 from 2-items set L2 and build 3-items set L3 form of candidate 3-items set C3. The candidate 3-element C3 are {1 2 3}, {2 3 4}, {1 2 5}, {2 3 4} and {2 3 5}. However, when forming the L2, items {1 3}, {1 5} and {4 5} of C2 in the remove. Thus, the candidate C3 items containing {1 3}, {1 5} or {4 5} will remove from C3. So only {2 3 5} which becomes item C3. List of items Set C3 can be seen in a table 9.

Table 9: Candidate 3-item set (C-3)

Itemset	Support
{2 3 5}	0.5525

Item set {2 3 5} is a frequent item. Furthermore, the algorithm stopped because it was no longer possible to obtain a higher item set or L4 = {}. So, items set L3 is only {2 3 5}. List of items Set L3 can be seen in table 10.

Table 10: Itemset (L-3)

Itemset	Support
{2,3,5}	0.5525

Stage3. This stage is the final stage. In this stage, we joint all k-frequent item-set (L_k). In this case, we joint L_1, L_2, and L_3. List of all set showed in table 11.

Table 11: All Itemset

Items Set	Support
{2,3,5}	0.5525
{1 2}	0.51
{1 4}	0.575
{2 3}	0.625
{2 4}	0.5225
{2 5}	0.5625
{3 5}	0.5525
{1}	0.6075
{2}	0.6975
{3}	0.665
{4}	0.6125
{5}	0.615

Using a pruning algorithm to avoid overlapping data, so we have a candidate CARS of class-1. See table 12.

Table 12: All Item set After Prune

Items Set	Support
{2,3,5}	0.5525

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

{1 2}	0.51
{1 4}	0.575
{2 4}	0.5225

Then calculate the confidence using equation 2. The confidence of item set {2 3 5} is counted as follows: item set {2 3 5} can formed from F_A = {3 5} and F_B = {2}.

So, $conf(F_A \Rightarrow F_B) = \frac{|f_A(x) \cap f_B(y)|}{|f_A(x)|} = \frac{0.5525}{0.5525} = 1$

Next, count confidence for another item set. So, we have table 13.

Table 13: Item set With Confedence

Itemset	Support	Confidence
{2,3,5}	0.55	1
{1 2}	0.51	0.84
{1 4}	0.58	0.95
{2 4}	0.52	0.85

Defined the confidence value threshold (*minconf*) and compare the confidence value of each possible rules with the *minconf*. Remove rules that have confidence value lower than *minconf*. If we defined the *minconf* = 0.90 or 90%, then the item set are {2,3,5} and {1,4}.

Table 14: All Item set

Items Set	Support	Confidence
{2,3,5}	0.55	1
{1 4}	0.58	0.95

Now, we have two items set {2 3 5} and {1 4}. Parameter 2, 3 and 5 have a relation to each other. Parameter 1 and 4 have relation to. Parameter 2, 3 and 5 have not relation with parameter 1 and 4. This information can be used for further analysis.

4. RESULT AND DISCUSSION

In this section, we implemented an FSAR algorithm for finding the relationship between parameter on student data set. The correspondent is a final year student of computer science at a college in Indonesia. In this case, we have 40 student data on important skills: hard skill and soft skill. There are 18 skill parameters that should be possessed by students to find a job. P1=Programming skill, P2=Reading Comprehension, P3=Critical Thinking, P4=Complex Problem Solving, P5=Active Listening, P6=Decision Making, P7=Time Management, P8=Writing, P9=System Analysis, P10=Active Learning, P11=Monitoring, P12=Ability to Communicate, P13=Social Perceptiveness, P14=Coordination, P15=Technology Design, P16=Persuasion, P17=Personal Resources, and P18=Team Work. The student will be proposed with a list of parameters that is related to the job requirements and they are required to give them a grade based on their qualifications and experiences. Every parameter may have scored Average=1, Good=2, Very Good=3 or Excellent = 4. The total score determines a student passed or not. If the total score exceeds the threshold, the student is passed. Unsuccessful students are required to take appropriate soft skill training.

Now, we will construct a fuzzy soft model for the data of student important skill. In this case, the process fuzzification can be done with simple normalization, just divide every score with the highest score.

Let C is a set of respondents (there are 40 students), so C = (C1, C2… C40) be an initial universe of object and P = (P1, P2… P18) is parameters which related to objects in C. Construct

$$F_P = \{ (x, f_p(x)) / \ x \in \text{P}, f_p(x) \in H(C) \}$$

Score of Pi (i = 1, 2 , …, 18) is membership of f_P where f_P is function which representing a mapping f_P:P→$H(C)$. The fuzzy soft set F_P represented in table 15.

So, F_P is a fuzzy soft set over C. Using minsupp = 0.67 and minconff = 1, after processed by FSAR algorithm, we will have 11 rules that listed in table 16.

Table 16: Item set, minsupp=0.67 and minconf=1

No	Rules	Support	Confidence
1	{P8 P11 P13}	0.717	1
2	{P8 P11 P18}	0.700	1
3	{P8 P12 P17}	0.693	1
4	{P8 P13 P18}	0.705	1
5	{P8 P17 P18}	0.681	1
6	{P8 P10 P11 P13}	0.681	1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7	{P8 P10 P11 P18}	0.675	1
8	{P8 P10 P13 P18}	0.681	1
9	{P8 P11 P13 P15}	0.686	1
10	{P8 P11 P15 P18}	0.681	1
11	{P8 P13 P15 P18}	0.686	1

In table 16 we see, P8 is exceeded in all of the rules, P8 is a significant parameter, it means P8 (writing skill) is very important. Let see a rule no 8, rules {P8 P10 P13 P18} have support =0.681 (68.1%) and confidence =1 (100%). Based on association rule theory by Agrawal [1], it is mean: if the students have a good skill in writing, Active Learning, Social Perceptiveness, they must be good as teamwork. The same thing is obtained from rule no1 1 and 2, if the student has skill P8 and P11, they must have P13, P18 skill in sequence. Furthermore from rule no 3, the students who have P8 and P12 skills, they will have P17 skill. Meanwhile, from 9, the students who have P8, P11 and P13 they must have P15 skill and from rule no 10 obtained, the students who have P8, P11, and P15, they will have P18 skill. From the above analysis, we can be stated that if the students have P8, P10, P11 and P12 they will have skill P13, P15, P17, and P18. Thus, if we want our students to have skill P8, P10, P11, P12, P13, P15, P17 and P18 (8 types of skills), we simply create a program to develop 4 types of skills only P8, P10, P11 and P12 only.

5. CONCLUSION AND FUTURE WORK

This study has been successfully implementing the fuzzy soft association rules to find a relationship between the parameters used to measure skill that must be held for a student. The results showed that using the 67% minimum support and minimum confidence 100%, We found 11 rules formed by the 8 types of skills (P8, P10, P11, P12, P13, P15, P17, and P18) are mutually influential. From the 8 parameters, we successfully found 4 significant skills (P8, P10, P11, P12). This means, by building a training program for 4 types of Skill, students will be able to have 8 types of skills. Developing programs for 4 skills are more efficient than compiling programs for 8 skills. Meanwhile, for the type of skill that does not appear in Table 11 such as programming skill, it means that the skill is not related to each other and must be developed respectively.

As future work, we will focus to develop the FSAR mining is becoming an associative classifier based on fuzzy soft sets and implements it in various disciplines.

ACKNOWLEDGEMENT

This research is partly supported by Tier 1 Research Grant Scheme, Vot H099. The lab facility is sponsored by Telkom University.

REFERENCES:

[1] R. Agrawal, “Mining Association Rules between Sets of Items in Large Databases,” in *Proceedings of ACM SIGMOD International Conference on Management of Data*, 1993, no. May, pp. 207–216.

[2] Y. Haralambous and P. Lenca, "Text Classification Using Association Rules, Dependency Pruning, and Hyperonymization," in *Proceedings of DMNLP, Workshop at ECML/PKDD, France*, 2014.

[3] S. M. Kamruzzaman, F. Haider, and A. R. Hasan, “Text Classification using Association Rule with a Hybrid Concept of Naive Bayes Classifier and Genetic Algorithm.”

[4] P. Karthik, M. Saurabh, and U. Chandrasekhar, “Classification of text documents using association rule mining with critical relative support based pruning,” *2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016*, pp. 1219–1223, 2016.

[5] A. R. Kulkarni, V. Tokekar, and P. Kulkarni, “Identifying context of text documents using Naïve Bayes classification and Apriori association rule mining,” *2012 CSI 6th Int. Conf. Softw. Eng. CONSEG 2012*, pp. 1–4, 2012.

[6] A. Martin, M. Manjula, and P. Venkatesan, “A Business Intelligence Model to Predict Bankruptcy using Financial Domain Ontology with Association Rule Mining Algorithm,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 8, no. 3, pp. 211–218, 2011.

[7] R. Hu, “Medical Data Mining Based on Association Rules,” *Comput. Inf. Sci.*, vol. 3, no. 4, pp. 104–108, 2010.

[8] Z. XU and R. Zhang, “Financial Revenue Analysis based on Association Rules Mining,” *2009 Second Asia-Pacific Conf. Comput. Intell. Ind. Appl.*, pp. 220–223, 2009.

[9] D. Molodtsov, "Soft Set T h e or y First R e s u l

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

t s," *Comput. Math. with Appl.*, vol. 37, pp. 19–31, 1999.

[10] A. R. Roy and P. K. Maji, "A fuzzy soft set theoretic approach to decision making problems," *Comput. Appl. Math.*, vol. 203, pp. 412–418, 2007.

[11] B. Handaga *et al.*, "An Algorithm for Classifying Numerical FSSC ," *Int. J. Fuzzy Syst. Appl.*, vol. 3, no. December 2012.

[12] A. Mangalampalli and V. Pudi, "Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets," in *IEEE International Conference on Fuzzy System*, 2009.

[13] Y. Jiang, H. Liu, Y. Tang, and Q. Chen, "Semantic decision making using ontology-based soft sets," *Math. Comput. Model.*, vol. 53, no. 5–6, pp. 1140–1149, 2011.

[14] D. Rohidin, N. A. Samsudin, and T. Herawan, "On mining association rules of real-valued items using fuzzy soft set," in *Advances in Intelligent Systems and Computing*, vol. 549 AISC, Springer, Charm, 2017, pp. 517–528.

[15] V. Babic and M. Slavkovic, "Soft and ard skills development : a current situation in Serbian companies," *Int. Conf. Manag. Knowl. Learn.*, pp. 407–414, 2011.

[16] L. Synder, D. Rupp, and T. G.C, "Personal Selection of Information Technology Worker: People, The Job, and Issues for Human Resource Management," in *Research in Personnel and Human Resources Management*, In J. J. M., Oxford: Elsevier., 2006.

[17] E. Rainsbury, D. Hodges, N. Burchell, and M. Lay, "Ranking workplace competencies: student and graduate perceptions," *Asia-Pacific J. Coop. Educ.*, vol. 3(2), pp. 8–18, 2002.

[18] D. R. Laker and J. L. Powell, "The Differences between Hard and Soft Skills and Their Relative Impact on Training Transfer," *Hum. Resour. Dev. Q.*, vol. 22(1), pp. 111–122, 2011.

[19] R. Agrawal, "Fast Algorithms for Mining Association Rules," in *Proceeding of the 20th International Conference on Very Large Data Bases (VLDB)*, 1994, pp. 487–499.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 15 Normalize Student Data On Important Skill

Student / Parameter	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
C1	0.5	0.75	0.5	0.75	0.75	0.5	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	1	0.5	0.75	1
C2	0.75	1	0.5	0.75	0.75	0.5	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.5	0.75	1	0.75	0.5
C3	1	0.75	0.25	0.75	0.75	0.75	0.25	0.75	0.75	1	1	1	0.5	0.5	0.75	0.5	0.5	0.75
C4	0.5	1	0.5	0.5	0.75	0.75	0.5	0.75	0.5	0.5	0.75	0.75	0.75	0.75	1	1	0.75	1
C5	0.5	0.75	1	0.75	0.75	0.5	0.5	0.75	0.5	1	0.75	0.75	0.75	0.5	0.75	0.75	0.75	0.75
C6	0.75	1	0.5	1	0.75	0.5	0.5	0.75	0.75	1	1	1	0.75	0.75	1	0.75	0.75	0.75
C7	0.75	0.75	1	0.25	0.75	0.75	1	1	0.5	0.5	1	0.75	1	0.5	0.75	0.5	1	0.75
C8	1	1	0.75	0.5	0.75	0.75	0.25	0.75	0.75	1	0.75	0.5	0.75	0.75	1	1	0.5	0.75
C9	0.75	1	0.5	0.75	0.75	0.75	0.5	0.75	1	0.5	0.75	0.75	0.5	0.75	0.75	0.5	0.75	1
C10	1	0.5	1	0.25	1	0.5	0.5	1	0.5	1	0.75	1	0.75	1	0.75	0.75	0.75	0.75
C11	0.75	1	0.5	0.75	1	0.75	0.5	0.75	1	1	0.75	0.75	0.75	0.5	1	0.5	1	0.75
C12	0.5	1	0.5	0.75	1	0.75	0.5	0.75	1	1	0.75	0.75	0.75	0.5	1	0.5	1	0.75
C13	0.5	1	0.5	0.5	0.5	0.5	0.5	1	0.5	1	0.75	0.5	1	0.5	1	0.5	0.5	1
C14	0.5	1	1	0.5	0.75	0.75	0.5	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.5	0.75	0.75
C15	0.75	1	0.5	0.75	1	0.75	0.5	0.75	0.75	0.75	0.75	0.75	0.5	0.5	0.75	0.5	0.75	0.75
C16	0.5	0.5	0.75	0.5	0.75	0.5	1	0.75	0.5	1	0.75	1	0.75	0.75	1	0.75	0.75	1
C17	0.5	0.75	0.25	0.75	0.25	0.75	0.5	0.75	0.75	0.75	0.75	0.75	1	0.5	0.75	0.5	0.75	0.75
C18	0.75	1	0.75	1	0.5	0.75	0.75	0.75	0.75	1	0.75	0.75	0.75	0.75	1	0.75	0.75	0.75
C19	0.75	0.75	0.5	0.75	1	0.5	0.75	0.75	0.75	0.5	0.75	0.75	0.75	1	0.75	0.75	0.75	0.75
C20	0.75	0.75	0.5	0.5	0.5	0.5	0.75	0.75	0.5	0.75	0.75	0.5	0.75	0.5	0.75	0.5	0.75	1
C21	0.5	0.5	0.5	0.5	1	0.5	0.75	0.75	1	1	0.5	0.75	0.75	0.75	0.75	0.75	0.75	0.75
C22	1	1	1	0.75	0.75	0.75	0.25	1	0.75	1	0.75	0.75	0.5	0.75	1	0.75	0.75	1
C23	0.75	1	0.5	0.75	0.75	1	0.5	0.75	0.5	1	0.75	1	1	0.75	1	0.75	0.75	0.75
C24	0.75	0.5	0.5	0.75	1	0.75	0.25	1	1	0.75	0.75	0.75	0.75	0.5	0.75	0.75	0.5	0.5
C25	0.5	1	0.5	0.25	1	0.75	0.75	1	0.5	0.5	1	0.75	1	1	0.5	0.75	0.75	0.5
C26	0.5	0.75	1	1	0.25	0.5	0.5	0.75	0.5	0.75	0.75	0.75	1	0.5	0.5	0.75	1	0.75
C27	0.5	0.75	1	0.5	0.75	1	0.5	0.75	0.75	1	0.75	0.75	0.75	0.5	0.75	0.75	0.75	1
C28	0.5	1	0.5	0.75	1	0.5	0.75	1	1	0.75	0.5	0.5	1	0.75	0.75	0.75	0.75	1
C29	0.25	0.5	1	0.75	1	0.5	0.5	1	0.5	0.75	0.75	0.75	0.75	0.5	1	0.5	0.5	1
C30	1	1	0.25	0.5	1	1	0.5	0.75	1	0.75	0.75	0.75	1	0.75	0.75	0.75	0.75	0.5
C31	1	1	1	0.25	0.75	0.75	0.75	1	0.5	1	0.25	1	0.75	0.5	0.75	0.75	1	0.75
C32	1	1	1	0.5	0.5	0.5	0.75	1	0.75	1	1	0.75	1	0.5	1	0.75	0.5	0.5
C33	0.25	0.75	0.25	0.75	0.5	0.75	0.5	1	0.75	0.75	0.75	1	0.75	0.75	0.75	0.5	1	0.75
C34	0.5	0.5	0.75	0.25	0.5	0.5	0.75	0.75	1	0.75	1	0.75	0.5	0.75	0.5	0.75	1	0.75
C35	0.75	0.75	0.5	0.75	1	0.5	1	0.75	0.75	0.75	0.75	1	0.75	0.5	0.75	0.75	0.75	0.75
C36	0.75	0.5	0.75	1	0.5	0.75	0.5	1	0.5	1	1	0.5	0.75	1	1	0.5	0.5	0.75
C37	0.5	0.5	1	0.5	0.5	0.5	0.75	1	0.75	0.75	0.75	0.75	1	0.75	0.75	0.75	0.5	0.75
C38	1	1	0.5	0.75	0.75	1	0.75	0.75	0.5	0.75	0.75	1	0.75	0.75	0.75	0.75	0.5	0.75
C39	0.75	1	0.5	0.75	0.5	0.75	0.75	1	1	0.75	0.75	0.75	0.75	0.5	1	0.75	0.75	0.75
C40	0.75	0.75	1	0.75	1	0.75	0.75	1	0.75	1	1	0.75	1	0.75	0.75	0.5	1	1