

Applying Rough Set Theory for Student Clustering on Assessment Datasets

Suhirman^{1,2}, Jasni Mohamad Zain²

¹University Technology Yogyakarta
Yogyakarta, Indonesia

²Faculty of Computer System and Software Engineering
Universiti Malaysia Pahang Gambang
Kuantan Pahang, Malaysia
suhirman@uty.ac.id, jasni@ump.edu.my

Abstract - Assessment is the last session of a lecture in college. There are several components that form the basis of calculations on the scoring end. Data clustering under rough set theory can be considered as a technique for data mining. A technique to select a clustering attribute based on rough set theory is presented. Max-Max Roughness is applied to select the clustering attributes. To find meaningful clusters from a dataset, clustering attribute is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent. Dataset is taken from a survey of 284 software engineering students. Data are taken on 5th semester students majoring in Information Engineering University of Technology of Yogyakarta Indonesia. Assessment consists of five components, namely two tasks, presentations, midterms and final exams. This assessment was conducted in 2011. The evaluation criteria used range from [0-100]. Student name, age, race, and attendance are not required in this assessment. In this study, we show how to determine the dominant attributes of a set of attributes of an assessment list by using the rough set theory (Max-Max roughness). The results obtained can potentially contribute to give a recommendation in awarding the final grade of a course more quickly and accurately.

IndexTerms - Assessment, Clustering, Rough set theory, Attributes

1. Introduction

Education is the foundation for achieving sustainable development. Concerning with the importance of this kind of education, the key aspect is needed on the measuring achievement levels in higher environmental education [1]. Higher education institutions are overwhelmed with huge amounts of information regarding student's enrollment, number of courses completed, achievement in each course, performance indicators and other data. This has led to an increasingly complex analysis process of the growing volume of data and to the incapability to take decisions regarding curricula reform and restructuring. On the other side, educational data mining is a growing field aiming at discovering knowledge from student's data in order to thoroughly understand the learning process and take appropriate actions to improve the student's performance and the quality of the courses delivery [2].

The techniques of grouping data of an object are similar, both to deal with certain conditions and in conditions of uncertainty. especially with high complexity, speeds and stability will be a problem in itself. The data collection can be described as vague and uncertain. Obtained clusters and applied queries do not necessarily have boundaries. Rough set theory was developed as a mathematical tool for dealing with vagueness and uncertainty. It is successfully applied in various tasks, the selection / attribute extraction, synthesis and classification rules, knowledge discovery, etc. Tolerance rough set model employing a tolerance relation is not an equivalence relationship in the original model of rough sets [3].

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification. To make learning process more effective, the educational systems deliver content adapted to specific user's needs. Adequate personalization requires the domain of learning to be described explicitly in a particular detail, involving relationships between knowledge elements referred to as concepts [4].

Clustering is a mostly unsupervised procedure and the majority of the clustering algorithms depend on certain assumptions in order to define the subgroups present in a data set. As a consequence, in most applications the resulting clustering scheme requires some sort of evaluation as regards to its validity.

2. Related Works

One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in examination, detection of abnormal values in the result sheets of the students, prediction about

students' performance and so on. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Data mining techniques in context of higher education by offering a data mining model for higher education system in the university [5].

The modest size dataset and well-defined problems are still rather hard to obtain meaningful and truly insightful results with a set of traditional data mining (DM) approaches and techniques including clustering, classification and association analysis [6]. The rough set theory can help solve uncertain problem well [7]. A system for analyzing student's results based on cluster analysis and uses standard statistical algorithms to arrange their scores data according to the level of their performance is described. A clustering algorithm is used for analyzing student's result data. The model was combined with the deterministic model to analyze the students' results of a private which is a good benchmark to monitor the progress of academic performance of students in higher Institution for the purpose of making an effective decision made by the academic planners [8].

Rough sets may provide representation of clusters, where it is possible for an object to belong to more than one cluster. This is of particular interest when buffer zones between clusters are a buffer zone required to diminish the clustering mistakes. The objects in such a buffer zone need a second look before eventually being assigned to a cluster [9]. Data clustering under rough set theory show how variable precision rough set model can be used to groups student in each study's anxiety. The results may potentially contribute to give a recommendation how to design intervention, to conduct a treatment in order to reduce anxiety and further to improve student's achievement [10].

The applicability data mining techniques are aimed to identify the main drivers of student satisfaction in education institutions [11]. The various data mining techniques like classification, clustering and relationship mining can be applied on educational data to predict the performance of a student in the examination and bring out betterment in his academic performance [12]. A computational method can efficiently assess the ability of students from of a learning environment capturing their problem solving processes [13]. An approach based on grammar guided genetic programming, which classifies students in order to predict their final grade based on features is extracted from logged data in a web based education system. This approach could be quite useful for early identification of students at risk, especially in very large classes, and allows the instructor to provide information about the most relevant activities to help students have a better chance to pass a course [14].

Predicting student failure at school has become a difficult challenge due to the high number of factors that can affect the low performance of students and the imbalanced nature of these types of datasets. A genetic programming algorithm and different data mining approaches are proposed for solving these problems using real data. To select the best attributes in order to resolve the problem of high dimensionality. Then, rebalancing of data and cost sensitive classification have been

applied in order to resolve the problem of classifying imbalanced data [15]. Research results show that there is a marginal difference, suggesting giving students scaffolding questions is less effective at promoting student learning than providing them delayed feedback [16].

The data may also be used to support and advise students in various ways, for the betterment of the student as well as the institute. Based on experience, the department claims to be able to distinguish the potentially successful students from the first year before the end semester. To do this in an early stage is important for the student as well as for the university, but the selection is only loosely based on assumption student similarities over the years. There is no thorough analysis. Data mining techniques may corroborate and improve the accuracy of prediction. Furthermore, data mining techniques may point out indicators of academic success that are missed until now [17]. Classification methods like Bayesian network, rule mining and decision trees can be used to extract the hidden knowledge about the students behavior. These methods can be applied on the educational data to identify the weak students and can also be used to predict the students behavior and performance in the examination [18].

The various data mining techniques can be applied to the set of educational data and what new explicit knowledge or models discover. The models are classified based on the type of techniques used, including prediction and description [19].

3. Rough Set Theory

In the 1980's, Pawlak introduced rough set theory to deal with this problem. [20]. Similar to rough set theory it is not an alternative to classical set theory but it is embedded in it. Concepts of the rough set theory are discussed for approximation, dependence and reduction of attributes, decision tables and decision rules. The applications of rough sets are discussed in pattern recognition, information processing, business and finance, industry, environment engineering, medical diagnosis and medical data analysis, system fault diagnosis and monitoring and intelligent control systems[21]. Rough set theory has attracted attention to many researchers and practitioners all over the world, who contributed essentially to its development and applications. The original goal of the rough set theory is induction of approximations of concepts. The idea consists of approximation of a subset by a pair of two precise concepts called the *lower approximation* and *upper approximation*. Intuitively, the lower approximation of a set consists of all elements that surely belong to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The difference of the upper approximation and the lower approximation is a *boundary region*. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. Motivation for rough set theory has come from the need to represent a subset of a universe in terms of equivalence classes of a partition of the universe. In

this chapter, the basic concept of rough set theory in terms of data is presented.

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table in the form of *attribute values*. By an *information system*, we mean a 4-tuple (quadruple), where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, V is the domain (value set) of attribute a , f is a total function such that, for every $u \in U$, $f(u, a)$ is called information (knowledge) function. An information system is also called a knowledge representation systems or an attribute-valued system and can be intuitively expressed in terms of an information table (refer to Table 1).

Table 1. An Information system

U/A	a_1	a_2	\dots	a_k	\dots	$a_{ A }$
u_1	$f(u_1, a_1)$	$f(u_1, a_2)$	\dots	$f(u_1, a_k)$	\dots	$f(u_1, a_{ A })$
u_2	$f(u_2, a_1)$	$f(u_2, a_2)$	\dots	$f(u_2, a_k)$	\dots	$f(u_2, a_{ A })$
u_3	$f(u_3, a_1)$	$f(u_3, a_2)$	\dots	$f(u_3, a_k)$	\dots	$f(u_3, a_{ A })$
\dots						
$u_{ U }$	$f(u_{ U }, a_1)$	$f(u_{ U }, a_2)$	\dots	$f(u_{ U }, a_k)$	\dots	$f(u_{ U }, a_{ A })$

This information is expressed by means of attributes used as descriptions of the objects. The data is treated from the perspective of set theory and none of the traditional assumptions of multivariate analysis are relevant. The Information system reveal all available knowledge about the object under review. In many applications, there is an outcome of classification that is known. This *a posteriori* knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A *decision system* is an information system of the form, where D is the set of *decision attributes* and C is the set of *condition attributes*. The elements of C are called *condition attributes*.

The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. Therefore, generally, we are unable to deal with single object. Nevertheless, we have to consider clusters of indiscernible objects. The following definition precisely defines the notion of indiscernibility relation between two objects.

Let $S = (U, A, V, f)$ be an information system and let B be any subset of A . Two elements $x, y \in U$ are said to be *B-indiscernible* (indiscernible by the set of attribute $B \subseteq A$ in S) if and only if $f(x, a) = f(y, a)$, for every $a \in B$.

Obviously, every subset of A induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute B , denoted by $IND(B)$, is an equivalence relation. It is well known that, an equivalence relation induces unique partition. The partition of U induced by $IND(B)$ in $S = (U, A, V, f)$ denoted by U/B and the equivalence class in the partition U/B containing $x \in U$, denoted by $[x]_B$.

The indiscernibility relation will be used to define set approximations that are the basic concepts of rough set theory. The notions of lower and upper approximations of a set can be defined as follows.

Let $S = (U, A, V, f)$ be an information system, let B be any subset of A and let X be any subset of U . The *B-lower approximation* of X , denoted by $\underline{B}(X)$ and *B-upper approximations* of X , denoted by $\overline{B}(X)$, respectively, are defined by

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\} \text{ and } \overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\}. \quad (1)$$

From Definition the following interpretations are obtained

- The *lower approximation* of a set X with respect to B is the set of all objects, which can be for *certain* classified as X using B (are certainly X in view of B).
- The *upper approximation* of a set X with respect to B is the set of all objects which can be *possibly* classified as X using B (are possibly X in view of B).

The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with

respect to $B \subseteq A$, denoted $\alpha_B(X)$ is measured by

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|} \quad (2)$$

where $|X|$ denotes the cardinality of X . For empty set \emptyset , it is defined that $\alpha_B(\emptyset) = 1$ [22]. Obviously, $0 \leq \alpha_B(X) \leq 1$. If X is a union of some equivalence classes of U , then $\alpha_B(X) = 1$. Thus, the set X is *crisp* (precise) with respect to B . And, if X is not a union of some equivalence classes of U , then $\alpha_B(X) < 1$. Thus, the set X is *rough* (imprecise) with respect to B [22]. This means that the higher of accuracy of approximation of any subset $X \subseteq U$ is the more precise (the less imprecise) of itself. It means that the concept "Decision" can be characterized partially employing attributes Analysis, Algebra and Statistics.

In this section, a technique for selecting a clustering attribute based on rough set theory is presented. This section, however, will be presenting the technique Max-Max Roughness to select the clustering attributes. To find meaningful clusters from a dataset, clustering attribute is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent. Model for selecting a clustering attribute is given in figure 1. While the steps to calculate the Max-Max Roughness shown in table 2.

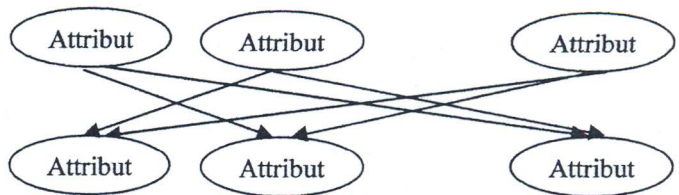


Figure 1. Model for selecting a clustering attribute

2	Each attribute in data set considered as a candidate attribute to partition
3	Determine equivalence classes of attribute-value pairs
4	Determine lower approximation of each equivalence classes in attribute a_i w.r.t. to attribute $a_j, i \neq j$
5	Determine upper approximation of each equivalence classes in attribute a_i w.r.t. to attribute $a_j, i \neq j$
6	Calculate roughness of each equivalence classes in attribute a_i w.r.t. to attribute $a_j, i \neq j$
7	Calculate mean roughness of attribute a_i w.r.t. to attribute $a_j, i \neq j$

Example, suppose data from student's values is given, as shown in table 3 There are 5 students with 5 attributes (assessment)

Table 3. A Student decision system

Student/ Assessment	A1	A2	A3	A4	A5
1	70	90	49	70	60
2	70	90	88	60	80
3	75	75	59	60	85
4	60	75	30	70	45
5	70	75	45	75	70

The following values are obtained from table 3

$U = \{1,2,3,4,5\}$; $A = \{A1,A2,A3,A4,A5,Grade\}$;

Where

$C = \{A1,A2,A3,A4,A5\}$; $D = \{Grade\}$;

$V_{A1} = \{60,70,75\}$; $V_{A2} = \{75,90\}$; $V_{A3} = \{30,45,49,59,88\}$;

$V_{A4} = \{60,70,75\}$;

$V_{A5} = \{45,60,70,80,85\}$; $V_{Grade} = \{A,B,C\}$

a. $S(A1=60)=\{4\}$, $S(A1=70)=\{1,2,5\}$, $S(A1=75)=\{3\}$

b. $S(A2=75)=\{3,4,5\}$, $S(A2=90)=\{1,2\}$

c. $S(A3=30)=\{4\}$, $S(A3=45)=\{5\}$, $S(A3=49)=\{1\}$,
 $S(A3=59)=\{3\}$, $S(A3=88)=\{2\}$

d. $S(A4=60)=\{2,3\}$, $S(A4=70)=\{1,4\}$, $S(A4=75)=\{5\}$

e. $S(A5=45)=\{4\}$, $S(A5=60)=\{1\}$, $S(A5=70)=\{5\}$,
 $S(A5=80)=\{2\}$, $S(A5=85)=\{3\}$

Calculating of roughness and mean roughness

First, determining of upper and lower approximations of singleton attribute with respect to other different singleton attribute. Then calculate the roughness and the average roughness of each attribute

Attribute A1

For attribute A1, it is clear that $|V(a_1)|=3$. The roughness and the mean roughness on A1 with respect to $Ai, i = 2,3,4,5$ is calculated as the following.

$$\underline{S}(A1 = 75) = \emptyset \text{ and } \overline{S}(A1 = 75) = \{3,4,5\}$$

Roughness

$$RA2(S | A1 = 60) = \frac{0}{3} = 0$$

$$RA2(S | A1 = 70) = \frac{2}{5} = 0.4$$

$$RA2(S | A1 = 75) = \frac{0}{3} = 0$$

Mean roughness

$$\text{Rough}_{A2}(A1) = \frac{0+0.4+0}{3} = 0.133$$

b. With respect to A3

The lower and upper approximations are

$$\underline{S}(A1 = 60) = \{4\} \text{ and } \overline{S}(A1 = 60) = \{4\}$$

$$\underline{S}(A1 = 70) = \{1,2,5\} \text{ and } \overline{S}(A1 = 70) = \{1,2,5\}$$

$$\underline{S}(A1 = 75) = \{3\} \text{ and } \overline{S}(A1 = 75) = \{3\}$$

Roughness

$$RA3(S | A1 = 60) = \frac{1}{1} = 1$$

$$RA3(S | A1 = 70) = \frac{3}{3} = 1$$

$$RA3(S | A1 = 75) = \frac{1}{1} = 1$$

Mean roughness

$$\text{Rough}_{A3}(A1) = \frac{1+1+1}{3} = 1$$

c. With respect to A4

The lower and upper approximations are

$$\underline{S}(A1 = 60) = \emptyset \text{ and } \overline{S}(A1 = 60) = \{1,4\}$$

$$\underline{S}(A1 = 70) = \{5\} \text{ and } \overline{S}(A1 = 70) = \{1,2,3,4,5\}$$

$$\underline{S}(A1 = 75) = \emptyset \text{ and } \overline{S}(A1 = 75) = \{2,3\}$$

Roughness

$$RA4(S | A1 = 60) = \frac{0}{2} = 0$$

$$RA4(S | A1 = 70) = \frac{1}{5} = 0.2$$

$$RA4(S | A1 = 75) = \frac{0}{2} = 0$$

Mean roughness

$$\text{Rough}_{A4}(A1) = \frac{0+0.2+0}{3} = 0.067$$

d. With respect to A5

The lower and upper approximations are

$$\underline{S}(A1 = 60) = \{4\} \text{ and } \overline{S}(A1 = 60) = \{4\}$$

$$\underline{S}(A1 = 70) = \{1,2,5\} \text{ and } \overline{S}(A1 = 70) = \{1,2,5\}$$

$$\underline{S}(A1 = 75) = \{3\} \text{ and } \overline{S}(A1 = 75) = \{3\}$$

Roughness

$$RA5(S|A1=60) = \frac{1}{1} = 1$$

$$RA5(S|A1=70) = \frac{3}{3} = 1$$

$$RA5(S|A1=75) = \frac{1}{1} = 1$$

Mean roughness

$$Rough_{A5}(A1) = \frac{1+1+1}{3} = 1$$

Likewise, The roughness and the mean roughness on A2, A3 and so on, thus obtained (table 4.)

Table 4. Mean Roughness Max-Max Roughness

	Mean Roughness				MMR
	a ₂	a ₃	a ₄	a ₅	
a ₁	0.133	1	0.067	1	1; 1; 0.133
a ₂	0.2	1	0.1	1	
a ₃	0.4	0	0.2	1	1; 0.4
a ₄	0.166	0	1	1	
a ₅	0.4	0	1	0.2	1; 0.4

Clustering result based on the Splitting attribute A2

--> { 1, 2 }
--> { 3, 4, 5 }

In this example where there are more than two attributes, the splitting is on the attribute value which has the overall maximum roughness versus the other attributes. The partition at this stage can be represented as a tree and is shown in figure 2.

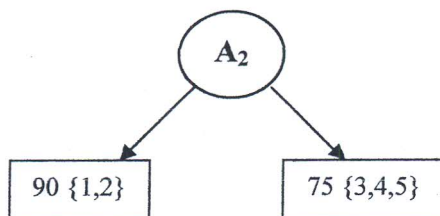


Figure 2. Result of clustering

4. Experimental Result

4.1. Datasets

Data were taken from the assessment of software engineering course at the informatics engineering program University Technology Yogyakarta Indonesia. The number of respondents was 284 students. Assessment consists of five components, namely the first task, the second task, midterms, the third task, and a final exam. This assessment is conducted in odd semester of academic year 2011/2012 of the fifth semester students. Evaluation criteria used range from [0-100]. The first task and the second task given to students individually to complete case studies provided by lecturers. The third task is a group of students made a paper to be

presented. Midterms done in the middle of the semester is done in writing, the final exams is given at the end of the semester. Both are done on a scheduled basis. students' name, age, race, and the force were not necessary in this assessment. From these data later in the process to give weight or a certain percentage to get the final value in the form of A for the highest value to the E to the lowest value. The rough set theory is to classify and determine the most dominant attributes of the five components. Sample of the data is shown in Table 5.

Table 5. List assessment of the software engineering course

Student / Assesment	A1	A2	A3	A4	A5
1	70	80	59	70	65
2	70	80	60	60	40
3	90	90	70	60	75
4	70	80	40	60	60
5	70	70	48	60	50

Description table :

A1: the first task

A4: midterm exam

A2: the second task

A5: final exam

A3: the third task

4.2. Result and Discussion

Table 6. Mean Roughness MMR

	Mean Roughness				MMR
	a ₂	a ₃	a ₄	a ₅	
a ₁	1	1	0,1	1	1
a ₂	0,3333	1	0,0667	1	
a ₃	0,2	0,4	0,2	1	1
a ₄	0,1	0,2	1	1	
a ₅	0,2	0,4	1	0,2	1

From table 6. Clustering result based on the Splitting attribute A1 : --> { 1,2,4,5 }
--> { 3 }

In software engineering course where there are more than two attributes, the splitting is on the attribute value which has the overall maximum roughness versus the other attributes. The partition at this stage can be represented as a tree and is shown in figure 3.

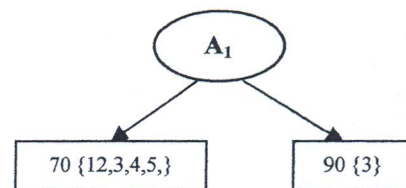


Figure 3. Result of clustering dataset

5. Conclusion

Rough set theory has been used as an attribute for the selection of a college student assessment. The approach described in this paper is Max-Max technique roughness (MMR). Data were obtained from the subjects of software engineering at the Department of Information Technology-University Technology Yogyakarta. The results indicate that the dominant attributes on other attributes can be specified, other attributes can be ignored, so that the process of assessment and provision of recommendations can be made more quickly. For future studies, the development of methods and software need to be better. Thus able to handle larger data and complex.

ACKNOWLEDGMENT

The authors would like to thank University Technology Yogyakarta and Universiti Malaysia Pahang for supporting this work.

REFERENCES

- [1] Petrovic, N., S. Isljamovic, and V. Jeremic, *A new concept for measuring achievement levels in higher environmental education*. Energy Education Science and Technology Part B: Social and Educational Studies. 5(3): p. 809-818,(2013).
- [2] Trandafili, E., et al. *Discovery and evaluation of student's profiles with machine learning*. in *5th Balkan Conference in Informatics*. Novi Sad,(2012).
- [3] Obadi, G., et al. *A Tolerance Rough Set Based Overlapping Clustering for the DBLP Data*. in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*.(2010).
- [4] Simko, M. and M. Bielikova. *Automatic Concept Relationships Discovery for an Adaptive E-course*. in *International Conference Educational Data Mining*. Cordoba, Spain,(2009).
- [5] Baradwaj, B.K. and S. Pal, *Mining Educational Data to Analyze Student Performance*. (IJACSA) International Journal of Advanced Computer Science and Applications. 2(6): p. 63-69,(2011).
- [6] Pechenizkiy, M., et al. *Mining the Student Assessment Data : Lessons Drawn from a Small Scale Case Study*. in *Int. Conf. Educational Data Mining*. Montreal, QC, Canada,(2008).
- [7] Chiu, D.Y., Y.C. Pan, and W.C. Chang. *Using Rough Set Theory to Construct E-Learning FAQ Retrieval Infrastructure*. in *IEEE Ubi-Media Computer Conference*. Lanzhou, China,(2008).
- [8] Oyelade, O.J., O.O. Oladipupo, and I.C. Obagbuwa, *Application of k-Means Clustering algorithm for prediction of Students' Academic Performance*. International Journal of Computer Science and Information Security. 7(1): p. 292-295,(2010).
- [9] Lingras, P. and G. Peters. *Applying Rough Set Concepts to Clustering*. in *Springer-Verlag London Limited*. London,(2012).
- [10] Yanto, I.T.R., et al., *Applying variable precision rough set model for clustering student suffering study's anxiety*. Expert Systems with Applications. 39: p. 452-459,(2012).
- [11] Dejaeger, K., et al., *Gaining insight into student satisfaction using comprehensible data mining techniques*. European Journal of Operational Research. 218: p. 548-562,(2012).
- [12] Kumar, S.A. and M.N. Vijayalakshmi. *Mining of student academic evaluation records in higher education in International Conference on Recent Advances in Computing and Software Systems*. Chennai,(2012).
- [13] Lee, Y.J., *Developing an efficient computational method that estimates the ability of students in a Web-based learning environment*. Computers & Education. 58: p. 579-589,(2012).
- [14] Zafra, A. and S. Ventura, *Multi-instance genetic programming for predicting student performance in web based educational environments*. Applied Soft Computing Journal. 12(8): p. 2693-2706,(2012).
- [15] Márquez-Vera, C., et al., *Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data*. Applied Intelligence: p. 1-16,(2012).
- [16] Feng, M., J.E. Beck, and N.T. Heffernan. *Using Learning Decomposition and Bootstrapping with Randomization to Compare the Impact of Different Educational Interventions on Learning*. in *International Conference Educational Data Mining*. Cordoba, Spain,(2009).
- [17] Dekker, G.W. *Predicting students drop out: a case study*. in *International Conference Educational Data Mining*. Cordoba, Spain,(2009).
- [18] Mishra, K.K., A. Tripathi, and A.K. Misra, *A performance evaluation tool for behavioral analysis of students*. Journal of Applied Sciences. 12(15): p. 1572-1579,(2012).
- [19] Delavari, N., S. Phon-amnuaisuk, and M. Beikzadeh, *Data mining application in higher learning institutions*. Informatics in Education Journal. 7(1): p. 31-54,(2008).
- [20] Pawlak, Z., *Rough Sets*. International Journal of Computer and Information Sciences. 11(5): p. 341-356,(1982).
- [21] Wu, C., et al., *The rough set theory and applications*. Engineering Computations: Int J for Computer-Aided Engineering. 21(5): p. 488-511,(2004).
- [22] Pawlak, Z. and A. Skowron, *Rudiments of rough sets*. Information Sciences. 177(1): p. 3-27,(2007).