International Journal of Pure and Applied Mathematics
Volume 119 No. 15 2018, 2239-2249
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

DECISION SUPPORT SYSTEM OF SCHOLARSHIP GRANTEE SELECTION USING DATA MINING

Eka Sugiyarti[1], Kamarul Azmi Jasmi[2], Bushrah Basiron[2], Miftachul Huda[2], Shankar K.[3], Andino Maseleno[1]

[1]Department of Information Systems, STMIK Pringsewu, Lampung, Indonesia
[2]Universiti Teknologi Malaysia, Malaysia
[3]School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, India

Corresponding author E-mail: andimaseleno@gmail.com

ABSTRACT
Education in a country life plays a very important role to ensure the survival of the state and the nation. External factors affect the failure of students in completing the field of study, especially the field of study in the senior high school level. The problems faced in getting the scholarship are some of the criteria that have been set in this course such as majors, parent income, parental responsibility, academic achievement, non-academic achievement. C4.5 algorithm is one method of data mining to predict the selection of scholarship grantee to senior high school student seen from external factors and internal students. The test results and analysis show that the Decision Tree C4.5 algorithm is accurately applied for prediction of final grades of senior high school students with 94.7368% accuracy.

Keywords: Data mining, Algorithm, C4.5, Scholarship grantee

I. INTRODUCTION

1.1 BACKGROUND

In the current technological development that grows very rapidly in the field of technology and education, many facilities that we get a scholarship which is a development program so that students can continue their study, scholarship assistance can be in the form of tuition, student learning assistance and an award for students who excel or for students who underprivileged to have a decision support system of scholarship grantee selection with data mining that can make it easier to get criteria of scholarship grantee.

According to research journal owned by Kustanto decision support system (DSS) web-based scholarship grantee selection is using data mining because data mining application can be used with large-scale database. The processed variables have dynamic weighting so that the application can determine the priority variable in the decision of education degree such as academic achievement improvement (ppa), the decision taken is more emphasized on the academic achievement and the scholarship for student learning aid (BBM) that is more emphasized on the economic conditions of prospective scholarship grantee [1].

According to research journal owned by Yoga Aditya Agassi, DSS scholarship is one of the development programs obtained by students to continue the study, both in the form of assistance costs, with various seminars and training activities. Furthermore it supports the cost of education, also an award for students because there is a class of scholarships earned on the achievement achieved by someone [2].

From two researches, the researcher wants to distinguish the research of DSS for the selection of scholarship grantee of one of the program in the form of cost assistance which is emphasized on the economy of the recipient candidate more in priority to the underprivileged party so that the student can continue the selection study. The scholarship recipient will use data mining because the processed variables have dynamic weighting.

The problems faced to get scholarships are some of the criteria that have been set in this study program such as majors, parent income, parental dependents, academic achievement, non academic achievement etc., therefore not all who enroll as a potential scholarship grantee may be accepted, only those who meet these criteria and conditions may receive scholarships, it is necessary to have a decision support system to assist in determining who is eligible to receive the scholarship.
Benefits and Purposes of DSS using data mining can be used with databases to facilitate the selection of scholarship grantee for the eligibility of scholarship grantee to obtain students for scholarships especially emphasized to underprivileged students and outstanding students, as with large-scale database applications can help make a scholarship recruitment decision.

1.2 PROBLEM FORMULATION

Based on the background above, there are several criteria in scholarship registration

1. Major
2. Academic achievement
3. Parents dependents
4. Parents income
5. Non academic achievement

1.3 RESEARCH

The purpose of this study is to obtain a Decision Support System for the eligibility of scholarship grantee for eligible students to receive scholarships.

II LITERATURE REVIEW

2.1 DEFINITION OF DECISION SUPPORT SYSTEM

According to Raymond McLeod, Jr. (1998) Decision Support System (DSS) is a system that provides the ability to solve problem and communications for semi-structured problems [3].
According to Bonczek (1980) Decision Supporting System (DSS) as a Computer-based System consisting of components, among others, component of language system, component of knowledge system and problem processing system (problem processing) interacting one on the other [3].

2.2 DEFINITION OF EDUCATION

In order to know the definition of education in policy perspective, we have had formal and operational formulation, as stated in Law No. 20 of 2003 on SISDIKNAS, namely:
Education is a deliberate and planned effort to create an atmosphere of learning atomsphere and learning process, so that learners actively develop their potential to have spiritual religion power, self-control, personality, intelligence, noble morals and skills needed by him/herself, nation society and nation.

2.3 DEFINITION OF SCHOLARSHIP

Scholarship is a form of financial assistance given to individuals intended to be used for the sustainability of educators who are required, scholarships can be provided by government agencies, companies or foundations.

2.4 DEFINITION OF DATA MINING

According to Urban in his book 'Decisions support systems and intelligent systems" data mining is a term used to describe the discovery of knowledge in the database, data mining is a process that uses statistical, mathematical, artificial and machine learning technique to extract and to identify pertinent information and related knowledge from a large data base [6].
Data mining is a process that uses stastistic, mathematical, artificial, and machine learning techniques to extract and identify useful information and related knowledge from various database [7,29].

2.5 TRAINING METHOD

Broadly training method data mining techniques are divided into two approaches namely:

1. Unsupervised learning, this method is applied without any training and without any teacher, the teacher here is the label of the data
2. Supervised learning, ie learning method with the practice and training. In this approach to find the function of decision, separator function or regression function, used some examples of data that have output or labels during the training process [6].

III. RESEARCH METHOD

3.1. OBSERVATION

This data collection is directly said through the field observation, in this research is done on the acceptance of scholarship along with the knowledge needs of scholarship. By doing this in the interview to the scholarship officials after making observations can be made framework concept based on information obtained.

3..2 INTERVIEW

That is data collection by conducting interviews directly with the teachers, staff or students to get input materials for the authors of this study. So the authors can find informations about school so that it can help the author to achieve the solution from (interview).

3.3 LITERATURE REVIEW

This literature study is a study conducted by looking for some sources or references either with books, articles or from other sources for reference system design and preparation of reports to be compiled.

3.4 C4.5 ALGHORITM

C4.5 Algorithm is an algorithm used to form decision tree, decision tree is a methodology and prediction that is very strongest and famous method of decision tree transforming very big fact into decision tree representing rule, rule can be easily understood with natural language, can also be expressed. Rules can be easily understood with natural language, and they can also be expressed in the form of database languages such as Stuctured Query language to search records in certain categories.

There are several stages to make a decision tree with C4.5 Algorithm :

1. Preparing training data, ordinary training data is retrieved from historical data that have occurred before and have been grouped previously to certain classes.
2. Specifiying the root in the tree root to be extracted from the selected attribute, by calculating the gain value of each attribute, the highest gain value which will be the first root, before calculating the gain value of the attribute, compute the entropy value used by the formula.

$$entropy\ (S) = \sum_{i=1}^{n} - pi.\log\ 2pi$$

Description

S = Custom set
n = The Number of S Partition
pi = Si proportion through S

3. Then calculate gain value use the formula :

Gain (S.A) = Entropy (s)

$$\sum_{i}^{n} \frac{|Si|}{|S|}$$

**Entropy (Si)*

Description :
S = custom set
A = feature
N = the number of A attribute partition
|Si| = Si proportion through..
|S| = The number of case of S

4. Repeat second stage until all record partitioned
5. the decision of tree partition process will stop when:
   a. all records in node N get the same class
   b. there are no attributes in the partitioned record
   c. there is no record in the empty branch [8]

3.5 CONFUSION MATRIX

Confusion matrix is a method that is usually used as an accuracy calculation of data mining concept. Information in confusion matrix is needed to determine the performance of the classification model, this summary information into a value is used to compare the performance of different models. This can be done by using performance metric.

classification	Predicted class	
	Class=Yes	Class=No
Class=Yes	a(true positif TP)	B(false negatif TN)
1.4 Class=No	**1.5 c(false positive-FP)**	**1.6 d(true negative-TN)**

1. ACCURACY

Calculation of this accuracy is by dividing the number of data and classification correctly with the total sample test data tested.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. PRECISION

Share the true positive value of true data divided by the number of true positive data and the false negative data (false negative)

$$Precision = \frac{TP}{TP + FP}$$

3. RECALL

Calculated by dividing the true positive (positive) data with the sum of the true positive data (true positive) and false negative (false negative) data

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F-MEASURE

This value is obtained from the calculation of the division of the multipication result of precision and recall with the sum of precision and recall then multiplied by two.

$$f - Measure = 2 * \frac{precision * recall}{precision + recall}$$

3.6 RESEARCH IDEA CONCEPT

In designing decision support system through several stages and processes so that later will get the results well and achieve the desired goal. some stages of this research can be illustrated in flowchart that can be seen in Figure 1 below.

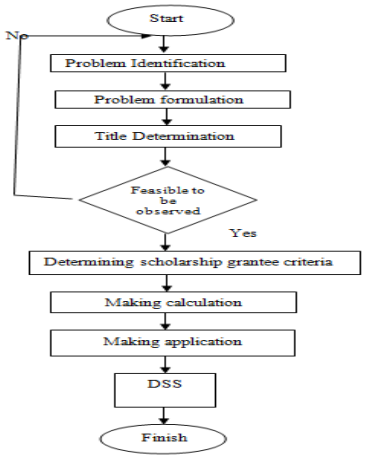

Figure 1. Flowchart of Scholarship Grantee Selection

Description

1. The user must identify the problem firstly, then formulate the next problem to determine the title.
2. After determining the title and then decide whether the data is complete feasible to be researched or not, otherwise it will be returned to identify the problem, if yes then the system will procees to determine the criteria for admission of prospective students.
3. After determining the criteria of scholarship grantee then make calculation.
4. After getting the input data or value then create the application
5. Further decision will be made for scholarship grantee.

IV. RESULTS AND DISCUSSION

4.1 MEASURED PARAMETER

The following table data majors are measured from the data majors with weighting value is shown in Table 1.

Table 1. Major data

No	Major
1	Natural Science
2	Social Science

The following is the parent income table as measured from the parent income table with the weighted value is shown in Table 2.

Table 2. Parents income

Group	Income	Category
Group 1	>=4000000	Upper class
Group 2	2100000-3900000	Middle class
Group 3	<=2000000	Lower class

The following table shows the number of parents dependent as measured by the number of dependents with the weighted value is shown in Table 3.

Table 3. The Number of Parents Dependent

Group	Dependent	Category
Group 1	<=2500000	Good
Group 2	>2500000	Bad

The following table data majors are measured from academic achievement by weighting its value is shown in Table 4. Table 5 shows academic achievement.

Table 4. Academic Achievement

Academic achievement	category
Rank 1,2,3	Good
Rank 4,5,6,7	Average
Rank >7	Poorly

Table 5. Non academic achievement

Non academic achievement	category
Not Achieving non academic achievement	Poorly
District level	Fair
Regency level	Average
Province level	Good
National level	Very Good

With preliminary figure for data collection for eligibility of scholarship acceptance.

C 4.5 Alghoritm measurement

Totalxp(juru	E(total*p(jurus	GAIN
0.5263158	0.961298027	0.020643
0.4349822		
0.1707954	0.890654132	0.091287
0.4018669		
0.3179918		
0.4018669	0.977358124	0.004583
0.5754912		
0.5110266	0.701007706	0.280933
0		
0.1899811		
0	0.362978787	0.618962
0		

With algorithm calculation can determine the gain.

With the c 4.5 alghoritm figure the gain calculation, to calculate the gain and to get the gain result.

Gain of major = 0.020643
Gain of parents income =0.091287
Gain of the number of parents dependents=0.004583
Gain of academic achievement=0.280933
Gain of non academic achievement =0.618962

From the calculation process of decision root tree determination above, it can be illustrated a decision tree.

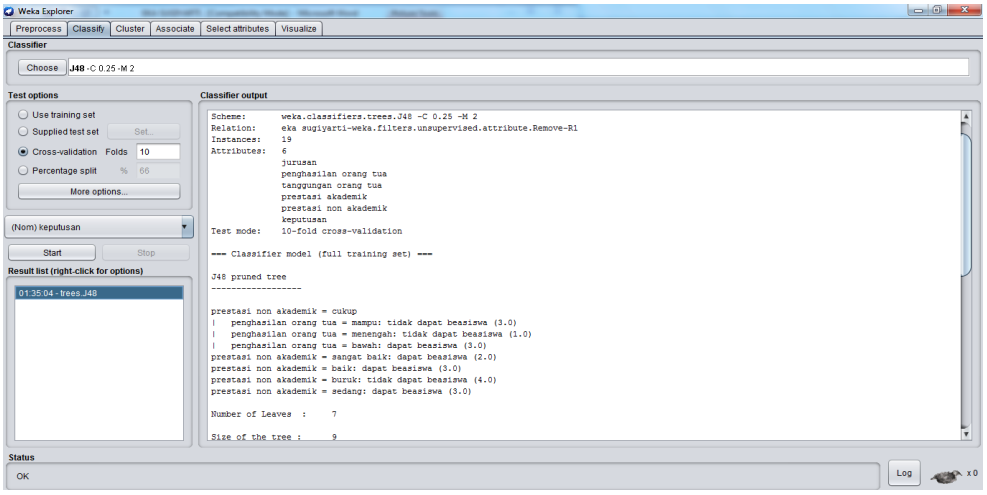

Figure 2. C4.5 algorithm calculation

From result of calculation of figure 2 will get process gain

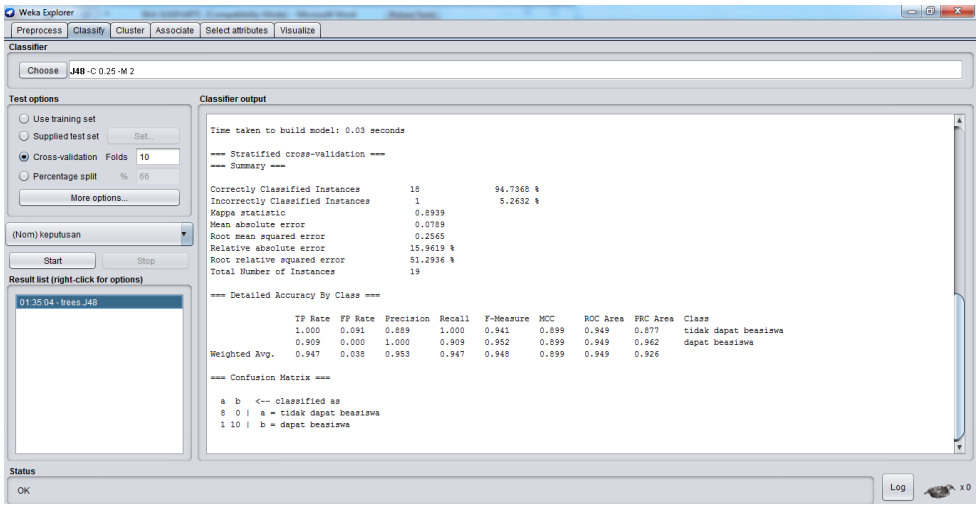

Figure 3. Process gain result

From the result of gain calculation will get gain to get decision tree as shown in Figure 4.

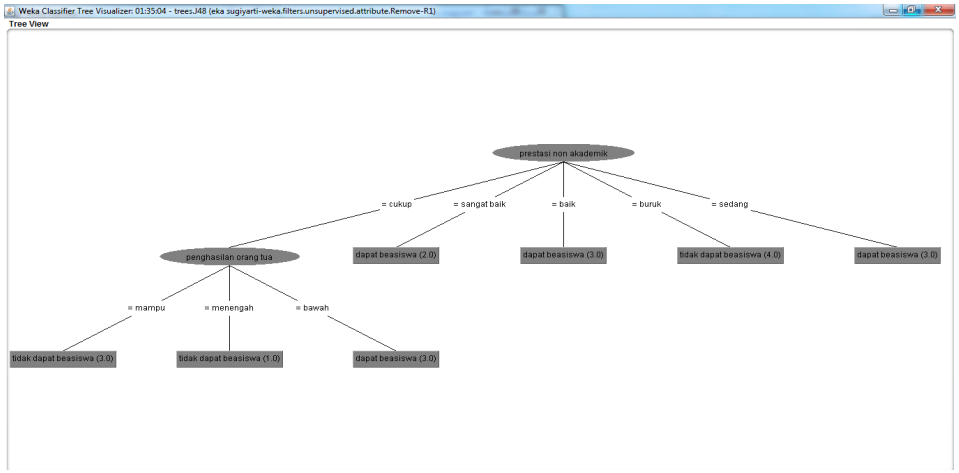

Figure 4. Decision tree model

From the results of gain calculation will get a decision tree.

With the decision tree model obtained the final result of the gain calculation that determines the dominant scholarship that is academic achievement

J48 pruned tree

Non academic achievement = fair
Parents income = upper class; not getting scholarship
Parents income = middle class: not getting scholarship (1.0)
Parents income = lowe class : getting scholarship (3.0)

Non academic achievement = very good: getting scholarship (2.0)
Non academic achievement = good : getting scholarship (3.0)
Non academic achievement = poorly ; not getting scholarship (4.0)
Non academic achievement = average: getting scholarship (3.0).

Number of Leaves: 7
Size of the tree: 9.

From the calculation results for the selection of scholarships for students using the C 4.5 algorithm can be summarized as follows:

1. If non academic achievement = fair
   And Parents income = upper class
   Then Not getting scholarship (3)

2. If non academic achievement = fair
   And parents income = middle class
   Then Not getting scholarship (1)

3. If academic achievement = fair And Parents income = Lower class
   Then getting scholarship (3)

   If non academic achievement = very good Then getting scholarship (2)

4. If non academic achievement = good Then getting scholarship (3).
5. If non academic achievement = poorly Then not getting scholarship (4)
6. If non academic achievement = average
   Then getting scholarship (3)

V. CONCLUSION AND SUGGESTION

5.1 CONCLUSION

The conclusions that can be illustrated from this journal is that a decision support system (DSS) can be generated for the eligibility of scholarship recipients for students who get a scholarship decision support system (DSS) with data mining method of C45 algorithm can help to better and make better effectiveness than before as well as in earnings reports obtained can be minimized in error.

5.2 SUGGESTION

Based on the testing process and the conclusion that has been done, then there are some suggestions in this research namely:

a. The number of data used are only 19 with the number of attributes are 9, so for better measurement results it is recommended to increase the number of larger data and the number of attributes.
b. Use other methods such as optimation, Genetic Algorithm (GA), Ant Colony Optimization and others. System aspect .

a. Improving the scholarship analysis system for determining the eligibility of scholarship potential grantee.
b. Measuring whether the model has been developed successfully or not, evaluation is used to measure the accuracy of the results achieved by the model.

REFERENCES

[1]. Khoirudin, 2008. *Decision Support System Providing Scholarship In Pt.Indomarco Prismatama Bandung Branch*, Faculty of Engineering and Computer Science, University Computer Indonesia, Bandung.
[2]. Turban, 2005 *Educational Research Methods*. Jakarta: Ghalia Indonesia.
[3]. Daihani, 2001 *Sensitivity Test Application For MADM Models Using SAW and TOPSIS Methods*, Yogyakarta.
[4]. Susilowati, Tri 2017 *decision support system for determining scholarship recipients in sman 1 bangunrejo using SAW STMIK Pringsewu* .
[5]. Khoirudin, 2008. *SNATI Decision Support System Determination of Eligibility of Eligible International School Pilot By Fuzzy Associative Memory Method*. Department of Informatics, Faculty of Industrial Technology, Islamic University of Indonesia.
[6]. Kusumadewi, 2007 *Diktat Artificial Intelligence Lecture, Department of Informatics, Faculty of Industrial Technology, Islamic University of Indonesia*. John Creswell (1996) *Fuzzy Multi-Attribute Decision Making (FUZZY MADM).* Yogyakarta: Graha Ilmu Publisher.
[7]. Rosli, M.R.B., Salamon, H.B., and Huda, M. (2018). Distribution Management of Zakat Fund: Recommended Proposal for Asnaf Riqab in Malaysia. *International Journal of Civil Engineering and Technology* 9(3), pp. 56–64.
[8]. Aminin, S., Huda, M., Ninsiana, W., and Dacholfany, M.I. (2018). Sustaining civic-based moral values: Insights from language learning and literature. *International Journal of Civil Engineering and Technology*. 9(4), 157-174.
[9]. Maseleno, A., Pardimin, Huda, M., Ramlan, Hehsan, A., Yusof, Y.M., Haron, Z., Ripin, M.N., Nor, N.H.M., and Junaidi, J. (2018a). Mathematical Theory of Evidence to Subject Expertise Diagnostic. *ICIC Express Letters*, 12 (4), 369 DOI: 10.24507/icicel.12.04.369.
[10]. Maseleno, A., Huda, M., Jasmi, K.A., Basiron, B., Mustari, I., Don, A.G., and Ahmad, R. (2018b). Hau-Kashyap approach for student's level of expertise. *Egyptian Informatics Journal*, doi.org/10.1016/j.eij.2018.04.001.
[11]. Maseleno, A., Huda, M., Siregar, M., Ahmad, R., Hehsan, A., Haron, Z., Ripin, M.N., Ihwani, S.S., and Jasmi, K.A. (2017). Combining the Previous Measure of Evidence to Educational Entrance Examination. *Journal of Artificial Intelligence* 10(3), 85-90.
[12]. Huda, M., & Teh, K. S. M. (2018). Empowering Professional and Ethical Competence on Reflective Teaching Practice in Digital Era. In Dikilitas, K., Mede, E., Atay D. (Eds). *Mentorship Strategies in Teacher Education* (pp. 136-152). Hershey, PA: IGI Global. doi: 10.4018/978-1-5225-4050-2.ch007
[13]. Huda, M., Teh, K.S.M., Nor, N.H.M., and Nor, M.B.M. (2018a). Transmitting Leadership Based Civic Responsibility: Insights from Service Learning. *International Journal of Ethics and Systems*, 34(1), 20-31.
[14]. Huda, M., Maseleno, A., Muhamad, N.H.N., Jasmi, K.A., Ahmad, A., Mustari, M.I., Basiron, B. (2018b). Big Data Emerging Technology: Insights into Innovative Environment for Online Learning Resources. *International Journal of Emerging Technologies in Learning* 13(1), 23-36. doi:10.3991/ijet.v13i01.6990
[15]. Huda, M., Maseleno, A., Teh, K.S.M., Don, A.G., Basiron, B., Jasmi, K.A., Mustari, M.I., Nasir, B.M., and Ahmad, R. (2018c). Understanding Modern Learning Environment (MLE) in Big Data Era. *International Journal of Emerging Technologies in Learning*. 13(5), 71-85. doi: 10.3991/ijet.v13i05.8042
[16]. Huda., M. & Sabani, N. (2018). Empowering Muslim Children's Spirituality in Malay Archipelago: Integration between National Philosophical Foundations and Tawakkul (Trust in God). *International Journal of Children's Spirituality*, 23(1), 81-94.
[17]. Huda, M., Jasmi, K. A., Mustari, M. I., Basiron, B., Mohamed, A. K., Embong, W., ... & Safar, J. (2017g). Innovative E-Therapy Service in Higher Education: Mobile Application Design. *International Journal of Interactive Mobile Technologies*, 11(4), 83-94.
[18]. Huda, M., Siregar, M., Ramlan, Rahman, S.K.A., Mat Teh, K.S., Said, H., Jamsari, E.A., Yacub, J., Dacholfany, M.I., & Ninsiana, W. (2017j). From Live Interaction to Virtual Interaction: An Exposure on the Moral Engagement in the Digital Era. *Journal of Theoretical and Applied Information Technology*, 95(19), 4964-4972.

[19]. Anshari, M., Almunawar, M. N., Shahrill, M., Wicaksono, D. K., & Huda, M. (2017). Smartphones usage in the classrooms: Learning aid or interference?. *Education and Information Technologies*, 22(6), 3063-3079.

[20]. Huda, M., Sabani, N., Shahrill, M., Jasmi, K. A., Basiron, B., & Mustari, M. I. (2017a). Empowering Learning Culture as Student Identity Construction in Higher Education. In A. Shahriar, & G. Syed (Eds.), *Student Culture and Identity in Higher Education* (pp. 160-179). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2551-6.ch010

[21]. Huda, M., Jasmi, K. A., Hehsan, A., Shahrill, M., Mustari, M. I., Basiron, B., & Gassama, S. K. (2017b). Empowering Children with Adaptive Technology Skills: Careful Engagement in the Digital Information Age. *International Electronic Journal of Elementary Education*, 9(3), 693-708.

[22]. Huda, M., Shahrill, M., Maseleno, A., Jasmi, K. A., Mustari, I., & and Basiron, B. (2017c). Exploring Adaptive Teaching Competencies in Big Data Era. *International Journal of Emerging Technologies in Learning*, 12(3), 68-83.

[23]. Huda, M., Jasmi, K. A., Basiran, B., Mustari, M. I. B., & Sabani, A. N. (2017d). Traditional Wisdom on Sustainable Learning: An Insightful View From Al-Zarnuji's Ta 'lim al-Muta 'allim. *SAGE Open*, 7(1), 1-8.

[24]. Huda, M., Jasmi, K. A., Embong, W. H., Safar, J., Mohamad, A. M., Mohamed, A. K., Muhamad, N. H., Alas, Y., & Rahman, S. K. (2017e). Nurturing Compassion-Based Empathy: Innovative Approach in Higher Education. In M. Badea, & M. Suditu (Eds.), *Violence Prevention and Safety Promotion in Higher Education Settings* (pp. 154-173). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2960-6.ch009

[25]. Huda, M., Jasmi, K. A., Alas, Y., Qodriah, S. L., Dacholfany, M. I., & Jamsari, E. A. (2017f). Empowering Civic Responsibility: Insights From Service Learning. In S. Burton (Ed.), *Engaged Scholarship and Civic Responsibility in Higher Education*(pp. 144-165). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-3649-9.ch007

[26]. Huda, M., Maseleno, A., Jasmi, K. A., Mustari, I., & Basiron, B. (2017k). Strengthening Interaction from Direct to Virtual Basis: Insights from Ethical and Professional Empowerment. *International Journal of Applied Engineering Research*, 12(17), 6901-6909.

[27]. Huda, M., Haron, Z., Ripin, M. N., Hehsan, A., & Yaacob, A. B. C. (2017l). Exploring Innovative Learning Environment (ILE): Big Data Era. *International Journal of Applied Engineering Research*, 12(17), 6678-6685.

[28]. Huda, M. (2018). Empowering Application Strategy in the Technology Adoption: Insights from Professional and Ethical Engagement. *Journal of Science and Technology Policy Management*. doi.org/10.1108/JSTPM-09-2017-0044.

[29]. Shankar, K. "Prediction of Most Risk Factors in Hepatitis Disease using Apriori Algorithm." RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES 8.5 (2017): 477-484.