Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

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

MINIMIZING STUDENT ATTRITION IN HIGHER LEARNING INSTITUTIONS IN MALAYSIA USING SUPPORT VECTOR MACHINE

[1]ANBUSELVAN SANGODIAH, [2]PRASHANTH BELEYA, [3]MANORANJITHAM MUNIANDY, [4]LIM EAN HENG, [5]CHARLES RAMENDRAN SPR

[1,4]Department of Information System, Faculty of Information and Communication Technology, University Tunku Abdul Rahman

[2]Department of Department of Physical and Mathematical Science, Faculty of Science, University Tunku Abdul Rahman

[3]Department of Computer Science, Faculty of Information and Communication Technology, University Tunku Abdul Rahman

[5]Department of Business, Faculty of Business and Finance, University Tunku Abdul Rahman

E-mail: [1]anbuselvan@utar.edu.my, [2]prashanthb@utar.edu.my, [3]manoranm@utar.edu.my, [4]ehlim@utar.edu.my, [5]charlesr@utar.edu.my

ABSTRACT

Attrition or better known as student dismissal or drop out from completing courses in higher learning institutions is prevalent in higher learning institutions in Malaysia and abroad. There are several reasons attributed to the attrition in the context of student in higher learning institutions. The degree of attrition varies from one institution to another and it is cause for concern as there will be a lot of wastage of resources of academic and administrative besides the adverse effect on the social aspect. In view of this, minimizing the attrition rate is of paramount importance in institutions. There have been numerous non-technical approaches to address the issue, but they have not been effective to predict at early stage the likelihood of students dropping out from higher learning institutions. Technical approach such as data mining has been used in predicting student attrition by some researchers in their past research work. However, not all prediction data mining techniques and other relevant and significant factors attributed to student attrition have been fully explored to address the issue. As of result this, this study will focus on using support vector machine model to predict probation status of student in which in most cases will lead to student's dismissal. It will also examine relevant and other factors that contribute to the attrition among students in Malaysia. The result of the study is appealing as the support vector machine model achieves a decent accuracy in prediction despite working on small size of data set. With all this in place, higher learning institutions in Malaysia can deploy the model in predicting probation status of student to minimize student attrition.

Keywords: *Attrition, Data Mining, Support Vector Machine, Classification Model, Educational Data Mining*

1. INTRODUCTION

Attrition becomes a normal scenario in many higher education institutions (HEIs) all over the world. The factors that lead to attrition among undergraduates and postgraduates in HEIs are varied from one to another taking into the consideration of geographical factors, ethnicity, education system of a country and etc [1]. Problem of attrition has been studied since 1980's where [2] and [3] have expressed the concern on attrition rate among students in HEIs. There are many studies have been carried out to find out the reasons which lead to attrition and also many investigations have been conducted focusing on overcome the attrition problem through various approaches including intervention strategies [4],[5].

The attrition rate among Malaysian students in both public and private HEIs is contributed by many factors. In private HEIs, most of attritions cases are due to financial issues since the tuition fee is very expensive, lack of facilities provided by the management and quality of teaching [6]. While the majority of students leave public HEIs because of two major reasons; (i) not able to continue their study in the programme that they enrolled due to lack of interest and (ii) failed in the examination [6]. These factors, which related to attrition need to

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

be handled carefully in order to solve the issue of attrition. Many strategies have been put forward to encounter this issue which mainly focuses on non-technical approaches that did not have much impact to solve the problems that lead to attrition. Therefore, an efficient mechanism using a technical approach needs to be identified and put forward that can guarantee to reduce the attrition rate and improve retention rate. Thus, in this study, the support vector machine model would be used to predict attrition rates. The reason that the model is chosen is due to its support for high dimensional space and sparse [7]. Besides that, it eliminates the need for feature selection hence making, it is easier to use. Also it is known for its superior performance compared to other data mining models in solving pattern recognition problem [8].

Every university and colleges have a strong aim to provide students different backgrounds with a conducive learning experience that could lead students to achieve success completing their study earning certificates, diploma, degree or postgraduate degrees. However, the aim could not reach every student where some of them might fail to complete their study due to many reasons. Attrition is disadvantageous for both students and HEIs. For students, fail to earn a degree will hamper their efforts to improve the socioeconomic status while for HEIs, attrition creates a major financial problem due to loss of income through tuition fees from students.

In Malaysia, the situation of students dropped out in their tertiary level education is alarming where, according to the latest statistics; out of 168000 college students who pursue their studies for certificates and diploma, 30000 would not graduate while , out of 100000 students who went for their degree programme, only 83000 able to finish the programme [9]. It means that 17.5% of total students who enrolled in tertiary education have dropped out in Malaysia. In [10] reported that a private university in Malaysia has an attrition rate exceeding 14% in just 6 months in the year 2012.

Universities outside Malaysia are also facing the attrition problem among the students. Griffith University in UK has an attrition rate of 21.2% in year 2011 [11] while [12] reported that in the year 2007, out of 32 Australian universities there were about average of 10.5% dropped out in each university [12]. The rate of attrition varies according to the field of study, in the US, students who enrolled in [13] related programs either change to another program or leaving colleges/universities without completing their study (48% - bachelor degree – and 69% - postgraduate level) between 2003 and 2009. In non-STEM programmes, both undergraduate and postgraduate suffered with an attrition rate of 56% between 2003 and 2009 [13]. Gardner [14] found that about 57% of doctoral students across disciplines have to leave HEIs without completing their degree. Petroff [15] cited that in the US, the attrition rate among PhD students is 50%.

HEIs all over the world are facing challenges in retaining number of students in their respective institutions where there is a need to identify factors that influencing students to leave higher education institutions before graduating. There are many factors from different aspects contribute to attrition; institutional environment, personal and financial problems are among the major reasons for students' attrition. Griffith University [11] indicated the following reasons for students' attrition which can be seen as universal reasons.

i. Personal difficulties – the most commonly given explanation for attrition, relating to health, finances, family, work, and difficulty fitting in or making friends.
ii. Academic difficulties – lack of academic preparedness, weak academic knowledge or specific study skills required to tackle the demands of the program; weak academic entry scores and low GPAs in first semester are all associated with greater attrition.
iii. Full time vs part-time status – part-time students are significantly less likely to continue into second year compared to full-time students.
iv. Making an uncertain or the wrong subject/program/university choice is linked to attrition. In some cases, this may reflect poor information provided prior to enrolment, or inadequate consideration of educational and career goals
v. Not being the university of first choice – a proportion of students leave one university to take up a more attractive opportunity at another institution if they are able to
vi. Loss of interest in the program or subject area
vii. Inability to manage time and workload demands and in consequence falling behind
viii. Dissatisfaction with the university experience, quality of curriculum or teaching.

Those numbers and reasons pertaining to the attrition rate among the students in HEIs showed that there is a need for a good mechanism that need to be developed and adopted by HEIs to overcome the problem of attrition in which this mechanism could hinder attrition from every angle of factors

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

that contribute to the rise of number of dropped out. Thus, in this investigation; there is a demand to holistically predict student attrition in HEIs in Malaysia and the support vector machine model will be used. Before a student is dismissed from a university, the student will undergo probation status in a row; hence there is strong relationship between attrition and probation status. In most of the cases in the context of attrition, students' dismissal will begin with probation status. Therefore in this study, predicting probation status of student which has close relationship with student attrition will be the focus.

This study is necessary as it is capable of identifying potential students in advance who will be on dropout list in the wake of growing number in student attrition in HEI in Malaysian and abroad.

2. LITERATURE REVIEW

Data mining (DM) is a computer-based information system (CBIS) [16] devoted to scan huge data repositories, generate information, and discover knowledge. The meaning of the traditional mining term biases the DM grounds. But, instead of searching natural minerals, the target is knowledge. DM pursues to find out data patterns, organize information of hidden relationships, structure association rules, estimate unknown items' values to classify objects, compose clusters of homogenous objects, and unveil many kinds of findings that are not easily produced by a classic CBIS. Thereby, DM outcomes represent a valuable support for decision making.

Concerning education, it is a novel DM application target for knowledge discovery, decisions-making, and recommendation [17]. Nowadays, the use of DM in the education arena is incipient and gives birth to the educational data mining (EDM) research field [18].

There are several data mining techniques or methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining [19]. As this study focuses on prediction through classification method, generally there are several models that are popular such as decision tree, neural network, support vector machine, Naïve Bayes and decision rule [20]. However in this study, support vector machine is used over other models to predict student attrition and the justification for using it in this study has been mentioned earlier.

There have been some past research work in predicting student attrition. A case study has been performed to predict electrical engineering students drop out in the Department of Electrical Engineering Eindhoven University of Technology. The experimental results show that rather simple and intuitive classifiers (decision trees) give a useful result with accuracies between 75 and 80% [21]. Certainly the research work is focused on decision tree model which are CART (SimpleCart) and C4.5 (J48), apart from other models such as a Bayesian classifier (BayesNet), a logistic model (SimpleLogistic), a rule-based learner (JRip) and the Random Forest (RandomForest).

The work by [22] uses three classification models which are logistic, decision tree and neural network in accordance to SEMMA methodology. Despite this research work primarily focused on comparing the accuracy between models in the context of student attrition, support vector machine model was not explored. Feature selection to determine contributing factors to student attrition was focused on non-academic factors. A total of 11 factors were used in the classifiers with the sample size of 715 students.

Similar research work has also been conducted at University of Science and Technology in Iran [23] where its focused classification models were regression and decision tree type of C5.0. The regressing is used to predict GPA while the decision tree is to predict students' dismissal. An accuracy of 88.5% in prediction has been recorded considering 13 factors consisting of selected academic and non-academic in predicting student attrition.

The importance of student dropout prediction using classification model is also evident in the research work by [24]. The work uses decision tree models which are CART and CHAID to identify the most contributing factors to student attrition and subsequently use those factors for prediction. An accuracy of 38.10% was recorded using CHAID while CART was 29% respectively. The work also ranked the identified factors to attrition which are financial, age, gender and disabilities using P-value and Chi-square and results show that only 3 factors (financial, age, gender) are significant to student attrition.

Research work on predicting drop out or attrition of learners not only confined to tertiary educations but also in schools as evidenced in [25]. This work focuses on induction rules and decision trees to make predictions. In order to improve the accuracy

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

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

in predicting which students might fail or dropout by first, all the available attributes are taken into consideration followed by selecting the best attributes and finally, rebalancing data and using cost sensitive classification. Two different experiments were conducted with 77 and 15 attributes respectively. Results indicate that the accuracy does not differ much between the two experiments. The size of data set is 670 students.

Other related work uses [26] prediction model particularly, artificial neural networks to predict the students' enrollment in a specific field at the end of the high school. This will identify appropriate field for the students for them to succeed in the field. Besides that, classification models have also been used to identify demographic, educational, and economic factors associated with a typically long time between doctoral program admission and degree completion [27].

The use of classification models such as neural network, decision tree and Naïve Bayes has also been popular in the education field to predict students' behavior as evidenced in these research work [28], [29].

The ability to predict learners' progression in tertiary education using a variety of classification models has been the main focus in the work by [30]. In the study, classification models such as Naïve Bayes (NB), unpruned Decision Tree5 (DT), Logistic Regression, Support Vector Machine using an Anova kernal function (SVM), Neural Network (NN) and k-Nearest Neighbour (k-NN) were used to compare the performance among the models in predicting students at risk of failing in the first year of study.

In a nutshell, the objective of this study is to focus on using support vector machine model with high number of vectors or predictors to observe the accuracy of student falling into the category of probation status in which early detection of it can minimize student attrition in general. This study is an extension of work by [31].

3. METHODOLOGY

As data mining is a process of discovering various models, summaries, and derived values from a given collection of data [32]. It is only appropriate that data that needs to be analyzed using data mining models should be accurate and reliable. The degree of accuracy of results obtained from data mining models is directly dependent on the accuracy of data, thus a data mining process or methodology is necessary. There are several well-known data mining methodologies and for this study Knowledge Discovery from Data (KDD) methodology is chosen. KDD is a process that provides, in such way, this knowledge. KDD [33],[34] is an interactive and iterative process [35]. It is a multi-stages process. According to [33] KDD process has 8 phases: problem formulation, data retrieval, data selection, data cleaning, data transformation, data mining, patterns evaluation, and knowledge integration and use as shown in Figure 1.

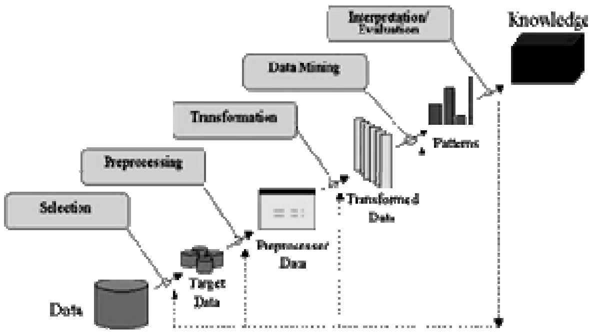

Figure 1: KDD

In this study, selection phase will involve in selecting relevant and important factors or attributes attributed to the probation status of student. Certainly a high number of vectors or attributes will be tested in this study. The selected attributes are expected to be academic and non-academic and the following proposed attributes will be considered.

Attributes/Dimensions: age, gender, race, occupation type of guardians, joint gross income of guardians, type of financial assistance, type of course, current GPA, current CGPA, GPA of first year of study, probation status currently, past history of probation status, entry qualification to university, qualification result, type of stream in secondary school, level of involvement in sports, level of involvement in social media, interest level in the current course, choice ranking of current course, ability of managing time, status of part time job, choice to change current course and medical status

Later in the preprocessing and transforming stages, there are inherent challenges in this study as the support vector model only supports input or data in the form of numerical and binary and most of the selected attributes are categorical based data. These attributes whose values are category based must be

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

converted into numeric or binary forms before they can be analyzed using support vector machine.

A simple technique can be deployed to convert a categorical attribute into binary form [36]. If there are m categorical values, then each original value can be uniquely assigned to an integer in the interval $[0, \; m-1]$. Thereafter, the each m integers can be converted to a binary number using n = $\log_2(m)$.

Example: parent's occupation attribute; assume that there are three possible category values which are professional, labor, others.

Categorical value	Integer value	x_1	x_2
Professional	0	0	0
Labor	1	0	1
Others	2	1	0

Besides binarization process as shown above, some attributes particularly income attribute though its value is expressed in numerical form but it has the potential to influence the final outcome as large values will take precedence over other attributes. Therefore, such attribute, its value will be normalized or scaled to prevent bias over other attributes.

Example: income attribute, assume the value is 20,000 and it will be normalized to a value between 0 to 1.

The next phase is about selecting a classification model. In this study, support vector machine model is chosen. Support Vector Machines, which were developed by [37] perform binary classification, i.e., separate a set of training vectors for two different classes (x_1, y_1), (x_2, y_2), ... ,(x_m, y_m), where $x_i \in R_d$ denotes vectors in a d-dimensional feature space and $y_i \in \{-1, +1\}$ is a class label. In this study, there will be two class labels which are likely in probation status and non-probation status. The strength of SVM lies in decision boundary of a linear classifier which separate class labels and it can be written in the following form:

$w \cdot x + b = 0$ where w and b are parameters of the model.

After putting some training data through SVM, the following graph will be observed as shown in Figure 2.

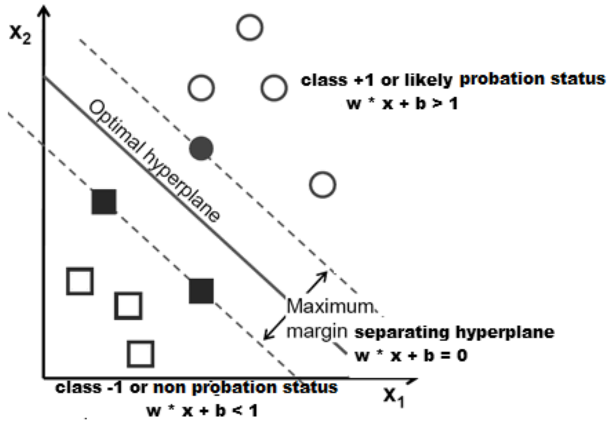

Figure 2: SVM decision boundary

Once the parameters of the decision boundary are found, a test instance z is classified as follows:

$$f(\mathbf{z}) = sign(w \cdot \mathbf{z} + b) = sign(\sum_{i=1}^{N} \lambda_i \; y_i \, x_i \cdot \mathbf{z} + b)$$

If $f(\mathbf{z}) > 1$, then the test instance is classified as a positive class or probation status (yes) class label otherwise, it is classified as a negative class or probation status (no) class label.

4. RESULTS AND DISCUSSION

4.1 Data Sets and Data Preprocessing

In this study, due to confidentiality issue, the data sets of actual number of students who have dropped out of HEI could not be obtained. Instead data sets of actual students who have been in probation status were obtained and used in this study. As mentioned earlier, there is a close relationship between students being in probation status and students dismissal. Students who have been in probation status multiple times are more likely to be dropped out of HEI. Therefore, only students who have been in probation status more than once have been selected.

A total of 150 students from the same university form the data set for this study. The attributes that are considered to be contributing factor to the student being in a probation status are as follows:

Attributes/Dimensions

age, gender, race, occupation type of guardians, joint gross income of guardians, type of financial assistance, type of course, current GPA, current

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

CGPA, GPA of first year of study, probation status currently, past history of probation status, entry qualification to university, qualification result, type of stream in secondary school, level of involvement in sports, level of involvement in social media, interest level in the current course, choice ranking of current course, ability of managing time, status of part time job, choice to change current course and medical status

The class label is probation status and out of 150 records of students, 21 records are students who are in probation status while the rest (129) are students who are not in probation status. As the latter (the number of records whose class label value is 'no' dominates almost the entire record), this has resulted in class imbalance. It must be resolved as most of the classifiers including support vector machine are biased towards the major classes and hence show very poor classification rates on minor classes [38]. It is also possible that classifier predicts everything as major class and ignores the minor class. Various techniques have been proposed to solve the problems associated with class imbalance [39], which divided into three basic categories, the algorithmic approach, data-preprocessing and feature selection approach. In this study data preprocessing approach is adopted. In data-preprocessing technique sampling is applied on data in which either new samples are added or existing samples are removed. Process of adding new sample in existing is known as over-sampling and process of removing a sample known as under-sampling. Over sampling is applied on the minor class (probation status = yes), where the existing 21 records were duplicated to 42 in total. As some of the records (particularly records associated with the major class (probation status = no)) contain lots of missing data, they were removed from the entire record.

The other data preprocessing work is converting nominal data into its equivalent in binary form. Attributes such gender, race, occupation type of guardians and others whose values are in nominal form were converted into binary form as described in [36].

4.2 Modelling

In this study, the data mining tool that is used is Rapid Miner 5.3. Once the data preprocessing work has been completed, the modelling work comes into play. Support Vector Machine model is used in predicting whether a student will be in a probation status. The following figures demonstrate how the modelling work has been carried out using the Rapid Miner tool.

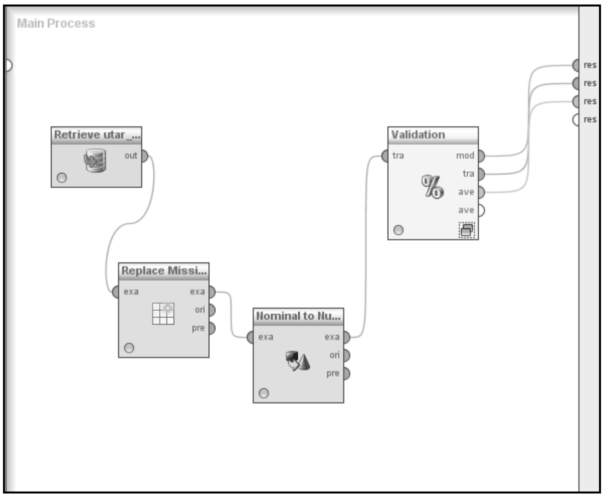

Figure 3: Overview of data processing task

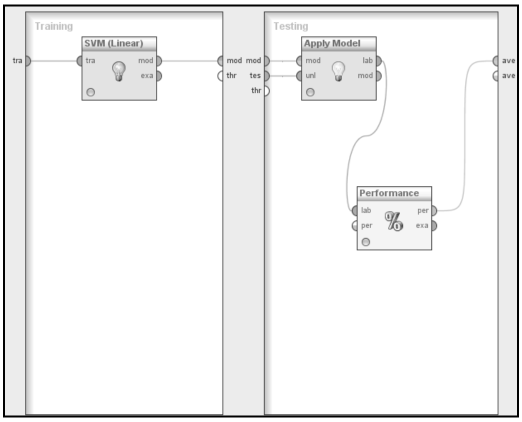

Figure 4: Overview of performance measure using Support Vector Machine

Validation operator as shown in Figure 3, is a nested operator where it contains other operators which are SVM, Apply Model and Performance (as shown in Figure 4). The validation operator uses cross validation method where it will partitioned the given data set into k subsets of equal size. Of the k subsets, a single subset is retained as the testing data set (i.e. input of the testing subprocess), and the remaining $k - 1$ subsets are used as training data set (i.e. input of the training subprocess). The cross-validation process is then repeated k times, with each of the k subsets used exactly once as the testing data. The value of k used in the validation operator is 10 and the sampling type is stratified.

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

And the type of Support Vector Machine model used is linear.

4.3 Results

The accuracy achieved with the data set using linear SVM appears to be encouraging, which is 89.84%. Also the class precision and the class recall for both class labels is appealing as well. The decent results as shown in Table 1 (confusion matrix) shows that even with small size of data set coupled with imbalanced proportionate ratio in class labels after over sampling has the discriminating power to predict probation status of student. Oversampling, removing incomplete records and proper handling of training and test data using cross validation method to a certain extent have contributed to the favorable results. Comparing with other methods such as decision tree and rule induction findings show that decent accuracy of 86.32% and 81.98% respectively can be achieved as shown in Table 2 and Table 3. In comparison to other methods it is noticeable that SVM works well with high dimensional data yet achieves a decent accuracy.

Table 1: Confusion Matrix with SVM

	True (NO)	True (YES)	class precision
pred NO	91	9	91.00%
pred YES	5	33	86.84%
class recall	94.79%	78.57%	

Table 2: Confusion Matrix with Decision Tree

	True (NO)	True (YES)	class precision
pred NO	82	5	94.25%
pred YES	14	37	72.55%
class recall	85.42%	88.10%	

Table 3: Confusion Matrix with Rule Induction

	True (NO)	True (YES)	class precision
pred NO	85	14	85.86%
pred YES	11	28	71.79%
class recall	88.54%	66.67%	

5. CONCLUSION

Apparently, this study concludes the potential use of classification model which is support vector machine to predict probation status of student in a holistic and accurate manner.

Certainly, higher education institutions would stand to benefit to the maximum if this study works on data sets consisting of actual number of students who have been dropped out of university. Despite the outcome of the study is favorable, working on larger data set is vital before using the SVM classifier for prediction to be put into practice in the real world context. Using different metrics apart from metrics such as accuracy, precision, recall to measure and detect the effect of imbalanced proportionate ratio in class labels is important.

Future work will involve larger data set involving SVM and to explore and take advantage to integrate physiological features in the study. The potential features might be extracted using Leonard Personality Inventory (LPI).

6. REFERENCES

[1] C. P. Veenstra, "A strategy for improving freshman college retention," *Journal for Quality & Participation*, vol. 31, no. 4, pp. 19-23, 2009.

[2] O. Lenning. (1980). *Retention and attrition: Evidence for action.* [Online]. Available: http://files.eric.ed.gov/fulltext/ED192661.pdf

[3] V. A. Stadtman. (1980). *Academic adaptations: Higher education prepares for the 1980s and 1990s*. San Francisco, California: Jossey-Bass, Inc. [Online]. Available: http://www.carnegiefoundation.org/publications/academic-adaptations-higher-education-prepares-1980s-and-1990s.

[4] S. Traci and K. Stefanie. (April 2012). "The best laid Plans: Examining the Conditions Under which a planning Intervention Improves learning and reduces attrition." *Journal of Applied Psychology*, doi: 10.1037/a0027977.

[5] K. Wilson. (2009). *The impact of institutional, programmatic and personal interventions on an effective and sustainable first-year student experience. In 12th First Year in Higher Education Conference 2009, Townsville.* Available:

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

http://www.fyhe.com.au/pa t_papers/ papers09/ ppts/Keithia_Wilson_paper.pdf

[6] W. Lesley, C. Julie and J. Sally. "Beyond the first-year experiences: the impact of attrition of student experience throughout undergraduate degree studies in six diverse universities," *Innovations in Education and Teaching International*, vol. 36, no. 3, pp. 331-35, 2011.

[7] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Proc. of the Tenth European Conference on Machine Learning (ECML '98)*, Lecture Notes in Computer Science, no. 1398, pp. 137–142, 1998.

[8] D. L. Lewis, Y. Yang, T. G. Rose and F. Li. "RCV1: A new benchmark collection for text categorization research," *Journal of Machine Learning Research*, 5: pp. 361-397, 2004.

[9] BorneoPost (2012). Available: http://www.theborneopost.com/2012/09/26/staggering-dropout-rate-before-spm/

[10] N. S. Govindarajo and D. Kumar. (2012). *How to Combat Attrition? Case Study on a Malaysian Educational Institution.* Available: http://cprenet.com/uploads/archive/IJBBS_12-1167.pdf

[11] Griffth University (2012). Available: http://www.griffith.edu.au/__data/assets/pdf_file/ 0006/ 419469/ Student-Retention-Strategy.pdf

[12] AUIDF (2007). Available: http://www.spre.com.au/download/ AUIDF Retention Results Findings.pdf

[13] STEM (2013). Available: http://nces.ed.gov/pubs2014/2014001rev.pdf

[14] S. K. Gardner. "Student and faculty attributions of attrition in high and low-completing doctoral programs in the United States," *Higher Education*, vol. 58, no. 1, pp. 97-112, 2008.

[15] L. Petroff. "The politics of graduate school," *National Social Science Journal*, vol. 35, no. 2, pp. 133-138, 2011.

[16] G. E. Vlahos, T. W. Ferratt and G. Knoepfle, "The use of computer-based information systems by German managers to support decision making," *Journal of Information & Management*, vol. 41, no. 6, pp. 763–779, 2004.

[17] C. Vialardi-Sacin, J. Bravo-Agapito, L. Shafti, and Ortigosa, "A. Recommendation in higher education using data mining techniques," *Proc. of the 2nd International Conference on Educational Data Mining*, pp. 190–199, 2009.

[18] A. Anjewierden, B. Kolloffel and C. Hulshof, " Towards educational data mining: using data mining methods for automated chat analysis to understand and support inquiry learning processes," *Proc. of the international workshop on applying data mining in e-Learning*, pp. 23–32, 2007.

[19] S. H. Liao, P. H. Chu, P. Y. Hsiao. "Data mining techniques and applications – A decade review from 2000 to 2011," *Journal of Expert Systems with Applications*, vol. 39, no. 12, pp. 11303-11311, 2012.

[20] P. A. Alejandro, "Educational data mining: A survey and a data mining-based analysis of recent works," *Journal of Expert Systems with Applications*, vol. 44, no. 4, pp. 1432, 2014.

[21] W. D. Gerben, P. Mykola and M. V. Jan. (2009). "Predicting Students Drop Out: A Case Study," *International Working Group on Educational Data Mining,* Available: http://eric.ed.gov/?id=ED539082.

[22] J. Mario, G. Željko, and M. Maja. **"**Student Dropout Analysis with Application of Data Mining Methods," *Management,* vol. 15, no. 1, pp. 31-46, 2010.

[23] N. Mahdi, M. Behrouz and V. Fereydoon. "Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining," *Proc. of 6th National and 3rd International conference of e-Learning and e-Teaching* (ICELET2012), IEEE, pp. 53-58, 2012.

[24] N. M. Mohammad, C. Linkon and S. K. Md. "Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country," *Proc. of IEEE/OSAIIAPR International Conference on Informatics, Electronics & Vision,* IEEE, pp. 113 -118, 2012.

[25] C. Marquez-Vera, C. R. Morales and S. V. Soto. "Predicting School Failure and Dropout by Using Data Mining Techniques,"

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

Tecnologias del Aprendizaje, IEEE Revista Iberoamericana de, vol. 8, no.1, pp. 7-14, 2013.

[26] M. M. Sayyed, A. A. Sayyede, “Prediction of Success or Fail of Students on Different Educational Majors at the End of the High School with Artificial Neural Networks Methods,” *International Journal of Innovation, Management and Technology,* vol. 4, no. 5, 2013.

[27] M. Gayle. Data Mining Analysis of the Effect of Educational, Demographic, and Economic Factors on Time from Doctoral Program Entry to Degree Completion in Education. *Dissertation: THE FLORIDA STATE UNIVERSITY COLLEGE OF EDUCATION,* 2006.

[28] Vasile, P.B. “Analysis and Predictions on Students’ Behavior Using Decision Trees in Weka Environment,” *Proceedings of the ITI 2007 29th Int. Conf. on Information Technology Interfaces,* IEEE, pp. 51-56, 2007.

[29] N. P. Oktariani. “Predicting Student Placement Class using Data Mining,” Proc. of *International Conference on Teaching, Assessment and Learning for Engineering (TALE),* IEEE, pp. 618-621, 2013.

[30] G. Gray, C. McGuinness and P. Owende. “An application of classification models to predict learner progression in tertiary education,” *Proc. In Advance Computing Conference (IACC),* IEEE, pp. 549-554, 2014.

[31] S. Anbuselvan and B. Balamuralithara, “Holistic Prediction of Student Attrition in Higher Learning Institutions in Malaysia Using Support Vector Machine Model,” *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE),* vol. 1, no 1, pp 29-35, 2014.

[32] K. Mehmed. *Data Mining: Concepts, Models, Methods and Algorithms Second Edition*. Wiley, 2011.

[33] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery Databases. In: Advances in knowledge discovery and data mining,” *American Association for Artificial Intelligence,* ISBN: 0-262-56097-6, pp. 1 – 34, 1996.

[34] CRISP-DM, Cross Industry Standard Process for Data Mining, available: http://www.crisp-dm.org, 2003.

[35] M. Ben Ayed, H. Ltifi, C. Kolski, and A. M. Alimi, “A User-centered Approach for the Design and Implementation of KDD-based DSS, DSS: A case Study in the Healthcare Domain,” *Decision Support Systems,* vol. 50, pp. 64–78, 2010.

[36] T. Pang-Ning, S. Michael, K. Vipin. *Introduction to Data mining*. Pearson, 2006.

[37] V. N. Vapnik. *Statistical learning theory*. New York: John Wiley and Sons, 1998.