

# Recommender System for Student Academic Performance based on Personality and Informal Learning

S.Dhanaraj  
Assistant Professor,  
RVS College of Engineering &  
Technology, Dindigul – 5.

A.Ramesh  
Assistant Professor,  
RVS College of Engineering &  
Technology, Dindigul – 5.

S.Suresh Kumar  
Associate Professor,  
RVS College of Engineering &  
Technology, Dindigul – 5.

## ABSTRACT

Educational Data Mining (EDM) mainly focuses on educational objectives like students' academic performance analysis based on personality and informal learning in formal learning environment. The primary objective is to identify the outcome of informal learning style (library and ICT) in formal learning environment. The secondary objectives are analyzing the students' personalities and purpose of resource utilization, identifying the resource usage level and etc. Eynseck Personality Questionnaire (EPQ) is used to classify the students' personality. Resource Utilization scale and Likert scale are used to measure the utilization of resource usage. Criterion Reference Model is used to classify the students' academic performance. Association rule is used to identify the frequent patterns among the set of attributes based on interesting measures. Multilayer perception technique provides the classification of confusion matrix result by applying cross-validation. This experiment can be used to improve the students' intellectual capability and understanding the subjects. This analysis can be used to predict the students' academic performance and the recommender system for students and management to improve the educative style of informal learning in formal learning environment and resource facilities.

**Keywords:** Personality, Impact of library and ICT, Academic performance, Association rule, Multilayer perception.

## 1. INTRODUCTION

Educational Data Mining is used to improve the students' academic performance and intellectual capability. EDM can solve many problems based on educational domain [16]. Data mining is non-trivial extraction of implicit, previously unknown and potentially useful information from large amounts of data. It is used to predict the future trends from the knowledge pattern [9]. Students are generally categorized into personalities like extraversion, neuroticism and psychoticism characters. The extraversion behaviorism is usually acted on the sociable, active and daring. The neuroticism behavior is called anxious, depressed and tense. The psychoticism behaviorism is generally abnormal characters like aggressive, egocentric and unsympathetic [11]. Informal learning factors are different from the traditional system. The informal learning style is embedded with the utilization of library and Information and

Communication Technology (ICT) resource. These resources are used to improve the students' intellectual development, information literacy, specialization knowledge and academic performance. The reading habit is to develop the memory and retention capability. The learning outcome is reflected in students' academic performance.

The performance is classified under the criterion reference model. To Collect the previous academic marks and classify the criterion below 50% is poor, 50 – 60 % is average, 60 – 70% is good, 70 – 80% is very good and above 80% is excellent. Student performance score is basically determined by the sum total of the continuous assessment and the examination scores [6]. Association rule is one of the techniques in data mining. Association rule includes identifying the frequent patterns based on the interesting measures. Frequent patterns are measured under minimum support and maximum confidence. The association rules are processed as follows: join step and prune step. Multilayer perception is one of the classification techniques in data mining. The classifications results are identified by confusion matrix. It has the training set, percentage split and cross-validation methods. It reveals that high accuracy and low error rate of results. It is classified under the category of personality, purpose of library and ICT usage and academic performance. MLP shows the TP Rate, FP Rate and ROC values. It is also used to check the correct and incorrect classified instances of the given input.

## 2. LITERATURE REVIEW

Nahyun Kwon et al. [11] examine the influence of personality traits on information competency. Analyze gender moderation of the relationship between personality traits and information competency. Students who are more conscientious, open to experience, and extroverted tend to greater information competency than students who are not. This research has also uncovered gender-specific personality traits affecting information competency. Agnes Ebi Maliki et al. [2] motivated the students about benefits of library usage and other resources. These benefits effect students' academic achievement and performance in reading ability level and intellectual development. These resources are conducive to improve the educational achievement.

The practice of reading skill is to enhance the learner's ability to perform mentally and physically.

Haruki Nagata et al. [8] explored the academic library functions that contribute to students educational outcomes. They observed how the students utilize the library resources and in what way the library resources are helpful to achieve their academic goals. Joe Frascotti et al. [10] provided a plan to improve the utilization of the library by the students at CCVC and help them to achieve academic success through increased the level of information literacy. The team developed a plan that consists of a set of recommendations explaining how the library use at the CCVC could be improved. They found some problems in library resources. The resources in the CCVC library were not currently adequate and did not serve the needs of the students. The location of the library is seventh floor so the library accessibility is a major problem. Book reading competition was recommended to motivate the students for increasing the library usage.

Nkoyo Edem et al. [13] motivated the students' library use and to update their knowledge and also assist their research work. They recommend aggressive user education and computerization of the library as well as acquisition of current information resources for proper and effective services. They also found reasons for satisfaction, dissatisfaction and expectations to become effective users of library resources. Recommendations are also made to the library management. The respondents indicated to strengthen their information retrieval process. Timothy Hebert et al. [17] survey results indicate the purpose of using library for both academic and social interest. A majority of students visits to participate in group work or to meet or study with friends. Majority of users are undergraduates and full-time students utilizing the library. It provides the level of library usage per week for each user. High respondents are using library as work on group projects.

Richard B. Lamptey et al. [15] examined the resources available and services offered by the KNUST Library. The vital role of library is dynamics of service delivery and speedy access of their resources. The members of library staff consider how the library could utilize the limited resources and achieve an effective level of service. Due to insufficient funds the KNUST library could not provide effective service to its users. They observed that the library was in dire need of fund for maintenance and further establishment. In addition to fund, the library needed a properly Library Bulletin containing the latest updates of resources which would help the users to interact with reference librarian. Emails should also be sent regularly to the lecturers informing them about the arrival of new journals which would help their research. Members of the general public who use the Library resources should be charged some fee as a way of generating some income to expand its facilities and improve its services. Adel Ben Youssef and Yves Punie et al. [1] [18] examined the relationship between ICT and students performance in higher education. Performance is evaluated on the basis of students' and teachers' characters, educational environment and the impact of ICT. The main objective is to improve the process of learning in higher education by offering new possibilities for learners and teachers and to analyze the possibilities to help their performance and achievement.

AreteValasidou et al. [4] their study found that students of political and social sciences were generally favorable to ICTs usage. Male students used ICT's more often than women.

Students' performance based on the use of computers at home tends to score higher than students with no access. This research helps to motivate the students to use ICT's in order to support their studies. ICTs knowledge plays a vital role in their entire working environment and the future gives them competitive advantage comparing to that are not familiar. Carlos E. and Naser Jamil et al. [5][12]'s objective of this research was to analyze the relations among the educative uses of ICT, level of skills in the technology handling and academic performance of Barineses university students. The technical skills are helpful to them in their future.

Chandra. E and Nandhini. K [3, 6] this paper identifies the reason for students' failure patterns and suggests to academic planners to improve the students academic successful achievement. The analysis reveals that more students fail due to their lack of ability. It also reveals some hidden patterns of students failed courses which could serve as base for academic planners in the curriculum restructuring and modification. It will help to improve the students' academic performance and reducing failure rate. The association rule found min. confidence should be of a higher percentage to be able to have more relevant and constructive rules.

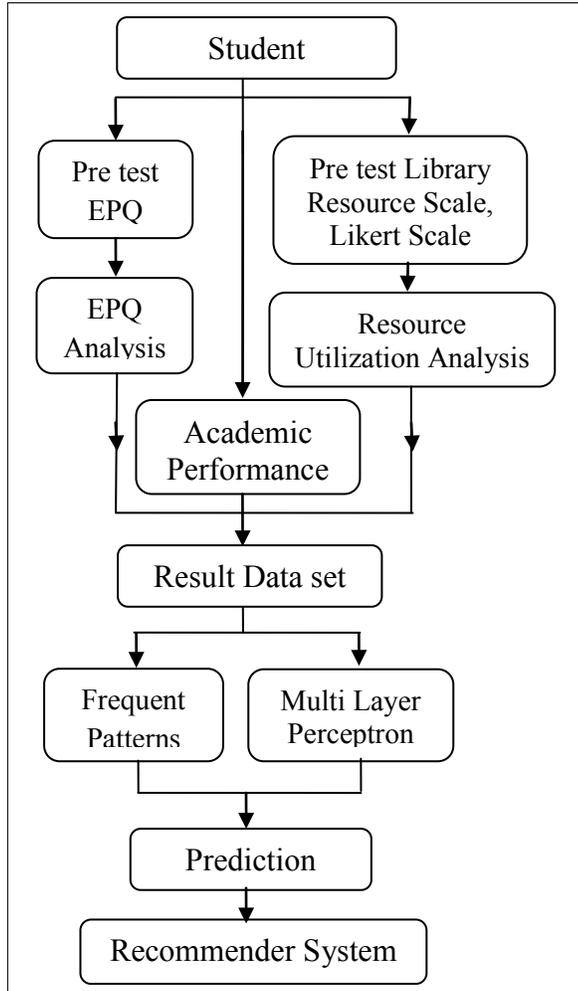
Deepa [7] this paper explains three data mining algorithms are applied on the task of classifying concrete compressive strength data set and the most accurate learning method is evaluated. Based on the result suggests that tree based modeling approach can effectively be used in predicting the compressive strength of high performance concrete. Oladokun et al. [14] this study has shown the potential of the artificial neural network for enhancing the effectiveness of a university admission system. This research is also recommended to non-engineering department itself.

### **3. PRP APPROACH**

The questionnaire is divided into three main areas. The first part of the questionnaire is demographic characteristics. The second part focuses on students' personality. The final part of the questionnaire depicts resource utilization. The Eysenck Personality Questionnaire (EPQ) is measured by E scale and N scale. The library and ICT questionnaires are categorized and measured by Resource Utilization scale and Likert scale (1-strongly disagree and 7- strongly agree). The Sample question can be taken, which is pretested.

The research is conducted from the post graduate college students. PRP approach can be divided into four phases: first phase describes pretest the Eynseck Personality Questionnaire and conducts the test among the students to categorize the personalities as extraversion, neuroticism and psychoticism. Second phase depicts on pretest the Resource Utilization Questionnaire and analyzes the resource utilization level, purpose of resource usage, user satisfaction of resource facilities and gender classification. Third phase is getting the students previous academic marks and categorize the performance based on Criterion Reference Model. The result dataset contains raw fact about the types of personality, informal learning factors (library and ICT) and academic marks. Result dataset is preprocessed and converted into normalize form to the input of data mining tools.

The fourth phase describes two approaches of data mining techniques: Association and Classification. First one is applying the association rule mining from the given dataset and associating between the set of attributes to identify the frequent patterns based on the interesting measures. The frequent patterns satisfy the minimum support and maximum confidence threshold. Association rules are containing the frequent item sets for the type of personality, resource usage level, purpose of resource usage and academic performance.



**Fig. 1 PRP Approach.**

The Fig. 1 shows the PRP approach and depicts the flow of data to perform the experiment. Multilayer Perceptron (MLP) is one of the classification techniques to classify various classes based on confusion matrix. Confusion matrix is used to identify the correct and incorrect classification values of given dataset. It gives high accuracy of results and low error rate of target and predicted values using cross-validation. Predict the students' academic performance based on the results using data mining techniques. To make a recommender system based on the prediction.

## 4. RESULTS AND DISCUSSION

### 4.1 Personality Analysis Result

The Alpha Miner tool found the frequent patterns from the given dataset, the extraversion character is used high level of library and ICT resources and they agree to these resources that are used for only study purposes and their academic performance is also very good. Weka tool produces confusion matrix classifier and it contains actual and predicted values of attributes in a given dataset.

**Table 1: Confusion matrix of personality based ICT users**

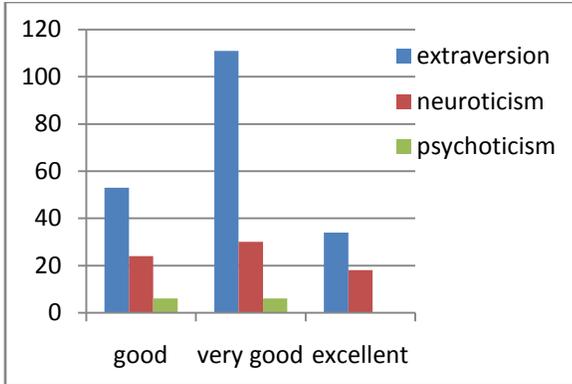
| CLASSIFIED AS    | A  | B   | C  |
|------------------|----|-----|----|
| A = NEUROTICISM  | 58 | 14  | 0  |
| B= EXTRAVERSION  | 7  | 191 | 0  |
| C = PSYCHOTICISM | 0  | 1   | 11 |

Table 1 shows the various classifications of personalities for ICT users, and the rate of accuracy defined as, the correct classification is (92.20%) and incorrect classification is (7.80%). Neuroticism, extraversion and psychoticism character of TP Rate is (0.806, 0.965, 0.917) and FP Rate is (0.033, 0.179, 0.0).

**Table 2: Confusion matrix of personality based library users**

| CLASSIFIED AS   | A  | B   | C |
|-----------------|----|-----|---|
| A= NEUROTICISM  | 68 | 4   | 0 |
| B=EXTRAVERSION  | 0  | 198 | 0 |
| C= PSYCHOTICISM | 2  | 2   | 8 |

Table 2 illustrates the extraversion, neuroticism and psychoticism characters of library users and their correct (97.16%) and incorrect (2.84%) classifications. F-Measure for neuroticism, extraversion and psychoticism is (0.958, 0.985, 0.8) and ROC Area is (0.946, 0.992, 0.846). The Fig. 2 shows that more students are of extraversion and neuroticism character and they are shining in their academic performance. It implies resource utilization of this type of character that is used only for studies. The psychoticism character has very low academic performance.



**Fig. 2 Personality based on academic performance.**

### 4.2 Resource Utilization Analysis Result

Table 3 gives the classifications of agree, disagree and neutral stand regarding the ICT usage for study purpose. It shows 69.86% of students agree ICT used for study purpose, 25.18% of students disagree. According to those who disagree, ICT is not used for study purpose and 2.13% of students are neutral. The remaining 2.83% of students are misclassifications.

**Table 3: Confusion matrix of ICT users for study purpose**

| CLASSIFIED AS | A   | B  | C |
|---------------|-----|----|---|
| A=ICT.A       | 197 | 5  | 0 |
| B=ICT.D       | 3   | 71 | 0 |
| C=ICT.N       | 0   | 0  | 6 |

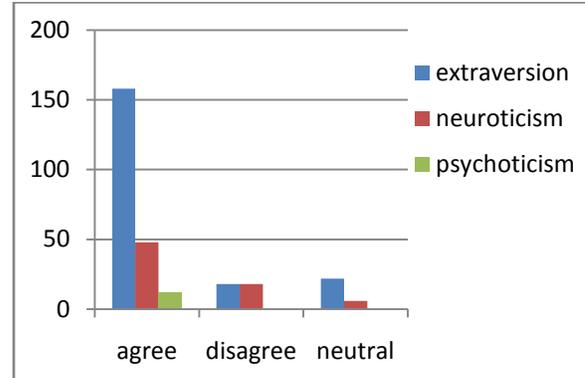
Table 4 shows that 75.89% of the students agree that the library is used for study purpose, 10.91% of students disagree and 6.38% of students are neutral.

**Table 4: Confusion matrix of library users for study purpose**

| CLASSIFIED AS | A  | B  | C   |
|---------------|----|----|-----|
| A=LIB.N       | 18 | 0  | 10  |
| B=LIB.D       | 3  | 31 | 2   |
| C=LIB.A       | 4  | 0  | 214 |

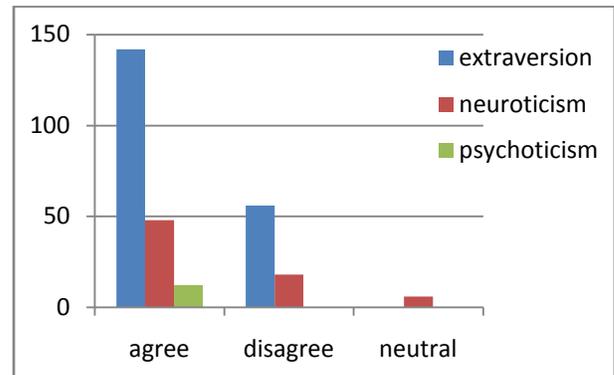
Table 5 gives agree, disagree and neutral stand towards various facilities in library and ICT resource. 63.05% of the students are satisfied with the resource facilities except for using the internet hours. Students are expecting more time to allow the ICT hours, to increase the resource from ICT and change the timing of resource utilization. It will help to improve the students'

knowledge. The Fig. 3 shows that extraversion, neuroticism and psychoticism characters agree for library usage only for study purpose. Low no.of extraversion and neuroticism characters disagree and neutral of library usage does not help academic purpose.



**Fig. 3 Library users based on personality.**

Fig. 4 shows that more no. of extraversion, neuroticism and psychoticism characters accept that ICT usage is helpful for improving the academic performance. Whereas less no. of extraversion and neuroticism characters are disagree with their opinion. It implies that ICT usage possesses both good and bad characteristics. But, it is general belief that whoever uses ICT for good purpose, he/she would certainly improve his knowledge.



**Fig. 4 ICT users based on personality.**

### 4.3 Students Performance Result

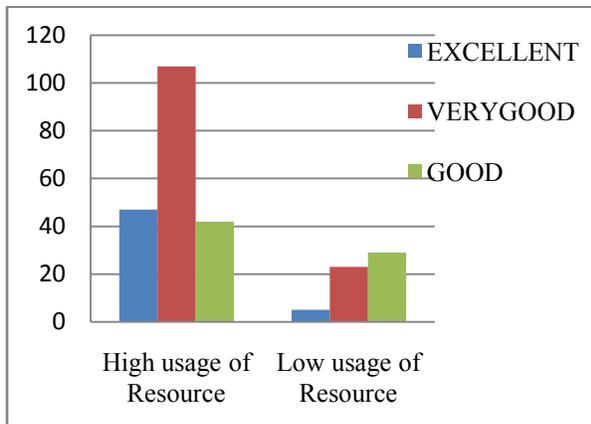
The students' academic performance is classified as excellent, very good and good based on the previous academic marks using Criterion Reference Model. Table 6 shows that more no. of students has good academic performance by utilizing the resources. This resource utilization will help to improve their knowledge and awareness of new technologies.

**Table 5: Satisfaction of library and ICT facilities**

| Library and ICT Facilities  | Agree  | Dis-Agree | Neutral |
|---|--------|-----------|---------|
| Are you satisfied with the information obtained from the library? | 76.60% | 6.38%     | 17.02%  |
| Is the library hours adequate?                                    | 24.82% | 24.82%    | 21.28%  |
| Is the library reading space adequate?                            | 6.38%  | 6.38%     | 9.93%   |
| Is the library collection sufficient?                             | 14.54% | 14.54%    | 19.15%  |
| Is the total number of books borrowing enough?                    | 39.72% | 8.51%     | 39.72%  |
| Is the xerox machine useful for library?                          | 8.16%  | 8.16%     | 12.06%  |
| Are the members of library staff friendly and willing to help?    | 53.90% | 27.31%    | 18.79%  |
| Are the computers in ICT sufficient?                              | 18.79% | 18.79%    | 29.43%  |
| Is the internet speed enough?                                     | 8.51%  | 8.51%     | 23.05%  |
| Is the internet hours adequate?                                   | 12.41% | 12.41%    | 43.26%  |

**Table 6: Confusion matrix academic performance**

| CLASSIFIED AS | A  | B  | C   |
|---------------|----|----|-----|
| A= EXCELLENT  | 30 | 9  | 13  |
| B= GOOD       | 5  | 59 | 19  |
| C= VERY GOOD  | 14 | 29 | 104 |



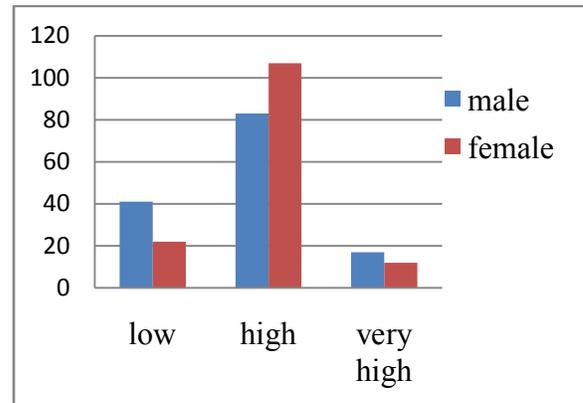
**Fig. 5 Academic performance based on informal learning.**

Fig. 5 shows informal learning of library and ICT usage helps to improve the students’ academic performance. Whoever uses high level of resource usage his academic performance is Excellent and Very good.

Whoever uses low level of resource usage they will have low academic performance.

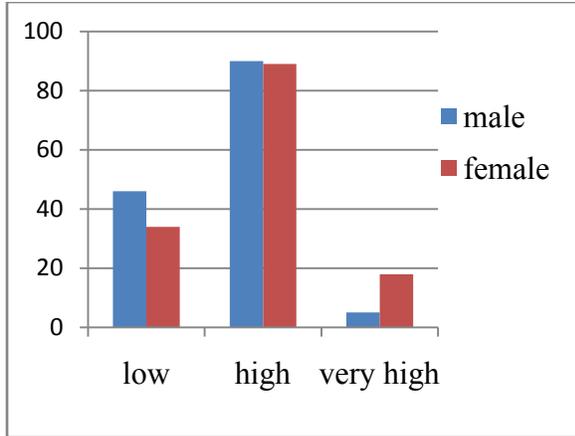
#### 4.4 Gender Analysis Result

Fig. 6 shows female students use library in high level and less no. of females use in a low level. It implies females are utilizing the library resource frequently. Male students are using low level of usage compared to females.



**Fig. 6 Genderwise library usage.**

Fig. 7 shows male and females are using ICT in a high level and less no. of males are using low level. It shows both genders are eager to utilize the ICT resource. Less no. of females mark low usage of ICT compared to male students.



**Fig. 7 Genderwise ICT usage.**

#### 4.5 Error Measures Analysis Result

Table 7 shows that error measures of various classifications based on the attributes of confusion matrix. It contains kappa statistics, mean absolute error, root mean squared error, relative absolute error and root relative squared error. It shows low error rate and high accuracy of results in multilayer perceptron technique.

#### 5. CONCLUSION

This experiment is useful for identifying the effect of informal learning and the academic performance and purpose of resource utilization based on personality and gender classification of resource usage. It analyzes the number of library and ICT users and their level of usage. It identifies the satisfaction of various resource facilities by the user. Association rule and multilayer perceptron techniques are used to identify the frequent patterns and classification results among the personality and resource utilization in student’s performance. Students’ academic performance based on the personality and resource utilization is predicted. A recommender system is made based on the prediction. Good character and high resource utilization will help to improve the academic performance and the management to recover and improve the resource facilities and increase the informal learners and the same is recommended to the student. Informal learning is helpful to improve their studies in a formal learning environment. This analysis is used to improve the students’ academic performance in educational domain. It gives a solution for reducing the students’ failure rate in the educational institution. Future scope of this research work is to include the aptitude test, technical test and proficiency of communication skill. Based on the results we can predict the students’ recruitment process. It is used to increase the rate of placement records, student admissions and popularity of the educational institution.

**Table 7: Error measures of MLP classification**

| Classification Factors   | Kappa statistic | Mean absolute error | Root mean squared error | Relative absolute error | Root relative squared error |
|--------------------------|-----------------|---------------------|-------------------------|-------------------------|-----------------------------|
| Academic performance     | 0.7547          | 0.0883              | 0.2094                  | 21.77%                  | 46.53%                      |
| Library utilization      | 0.7368          | 0.103               | 0.2272                  | 31.68%                  | 56.47%                      |
| Purpose of library usage | 0.8116          | 0.0494              | 0.166                   | 19.53%                  | 46.86%                      |
| ICT utilization          | 0.7849          | 0.0843              | 0.2158                  | 24.69%                  | 52.32%                      |
| Purpose of ICT usage     | 0.9326          | 0.0191              | 0.09                    | 6.82%                   | 24.10%                      |
| Personality              | 0.9336          | 0.0262              | 0.1314                  | 8.89%                   | 34.29%                      |

## 6. REFERENCES

- [1] Adel Ben Youssef and MounirDahmani, “The impact of ICT’s on students’ performance in Higher Education: Direct effects, Indirect effects and Organizational change”, 2010.
- [2] Agnes Ebi Maliki1 and Rachel D. Uche, “Students’ Background Variables and Utilization of Library Resources among Secondary School Students’ in Southern Senatorial District of Cross River State, Nigeria”, *Stud. Tribes Tribals*, 5(1): 21-23, 2007.
- [3] Agrawal. R, Imielinski. T and Swami. A, “Mining association rules between sets of items in large databases”, *Proceedings of the ACM SIGMOD Conference*, Washington, DC (1993).
- [4] AretiValasidou and DespoinaBousiou-Makridou, “The Impact of ICT’s In Education: The Case of University of Macedonia Students”, *Journal of Business Case Studies*, Vol 4, Num 3, 2008.
- [5] Carlos E. Godoy Rodríguez and Ezequiel Zamora, “Educative uses of ICT, technological skills and academic performance of the Venezuelan university students” *IJEDICT*, Vol. 2, Issue 4, pp. 2843, 2006.
- [6] Chandra. E and Nandhini. K, “Knowledge Mining from Student Data”, *European Journal of Scientific Research*, ISSN 1450-216X Vol.47 No.1, pp.156-163, 2010.
- [7] Deepa.C , SathiyaKumari.K and Pream Sudha.V ,“Prediction of the Compressive Strength of High Performance Concrete Mix using Tree Based Modeling”, *International Journal of Computer Applications (0975 – 8887)Volume 6– No.5, September 2010*
- [8] Haruki Nagata, Akira Toda and PaiviKytomaki, “Students’ Patterns of Library Use and Their Learning Outcomes”, 2008.
- [9] Jiawei Han and Micheline Kamber, “Data Mining: concepts and techniques”, Morgan Kaufmann Publishers, San Francisco, 2006.
- [10] Joe Frascotti, Jamie Levenseler, Collin Weingarten and Karl Wiegand, “Improving Library Use and Information Literacy at Caritas Charles Vath College”, 2007.
- [11] Nahyun Kwon and Hana Song, “Personality traits, gender, and information competency among college students”,*Malaysian Journal of Library & Information Science*, Vol. 16, no. 1, 87-107, April 2011.
- [12] NaserJamil Al-Zaidiyeen, Leong Lai Mei and Fong Soon Fook, “Teachers’ Attitudes and Levels of Technology Use in Classrooms: The Case of Jordan Schools”, *International Education Studies* Vol. 3, No. 2; May 2010.
- [13] Nkoyo Edem, Okon Ani and Jonathan A. Ocheibi, “Students’ perceived effectiveness in the use of library resources in Nigerian universities”, *Educational Research and Review* Vol. 4(6), pp. 322-326, June, 2009.
- [14] Oladokun.V.O, Adebajo.A.T, and Charles-Owaba.O.E, “Predicting Students’ Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course”, *The Pacific Journal of Science and Technology*, Volume 9. Number 1. May-June 2008.
- [15] Richard B. Lamptey, “Promoting Effective Use of Library Resources and Services at Kwame Nkrumah University of Science and Technology Library”, 2010.
- [16] Romero.C and Ventura.S, "Educational data Mining: A Survey from 1995 to 2005".*Expert Systems with Applications (33)*, 2007, pp.135-146.
- [17] Timothy Hebert and Robert Wolk, “Back to the Maxwell Library’s Future Student Library and Information Resource Usage”, *Proc ISECON 2006*.
- [18] Yves Punie, Dieter Zinnbauer and Marcelino Cabrera, “A Review of the Impact of ICT on Learning” Working Paper prepared for DG EAC, JRC 47246, October 2006.

## 7. AUTHORS PROFILE

**S. Dhanaraj** is working as an Assistant Professor in Department of Computer Applications, RVS College of Engineering & Technology, Dindigul, Tamil Nadu, India. He has 3 years of experience in teaching field and 1 year experience in his research area. His current area of research is Data Mining.

**A. Ramesh** is working as an Assistant Professor in Department of Computer Applications, RVS College of Engineering & Technology, Dindigul, Tamil Nadu, India. He has 1 year experience in his research area. His current area of research is Data Mining.

**S. Suresh Kumar** is working as an Associate Professor in Department of Computer Applications, RVS College of Engineering & Technology, Dindigul, Tamil Nadu, India. He has 13 years of experience in teaching field. His current area of research is Software Engineering and Data Mining.