Australian Journal of Basic and Applied Sciences, 9(27) August 2015, Pages: 433-441

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

An Analysis of Association and Classification Rules for Data Mining Algorithms with Student Dataset

[1]R. Priya and [2]Dr. Ananthi Sheshasaayee

[1]*Research Scholar, Department of MCA, Vels University, Pallavaram, Chennai, INDIA*
[2]*Head & Associate Professor Department of Computer Science Quaid-e-millath govt. college for women (autonomous), Chennai-2, INDIA*

ARTICLE INFO

Article history:
Received 23 June 2015
Accepted 25 August 2015
Available online 2 September 2015

Keywords:
Association Rule Data Mining,Apriori, FP-growth, DynFP growth, Student Stress, MATLAB

ABSTRACT

In recent years, the analysis and evaluation of students data set performance and retaining the standard of education is a very important problem in all the educational institutions. The most important goal of the paper is to analyze and evaluate the school and college students‟ performance by applying data mining classification algorithms in MATLAB tool. The data mining tool has been generally accepted as a decision making tool to facilitate better resource utilization in terms of students‟ performance. The compared algorithms are presented together with some experimental data that lead to the final conclusion and decisions.

To Cite This Article: R. Priya and Dr. Ananthi Sheshasaayee., An Analysis of Association and Classification Rules for Data Mining Algorithms with Student Dataset. ***Aust. J. Basic & Appl. Sci.,*** *9(27): 433-441, 2015*

INTRODUCTION

College life can be very stressful. Sometimes parents, faculty and others tend to idealize their college experience and remember it as that idyllic time when they had few worries or responsibilities. To students currently attending college, however, the process is often stressful and frustrating. Before condemning stress outright, we need to understand that stress is only harmful when it is excessive. Much of the stress that we all experience is helpful and stimulating. The challenges of life tend to be stressful and an attempt to avoid stress completely would lead to a rather boring existence. The problem comes when you experience too much stress.

Although some stress reactions are part of deeper and more serious emotional problems, many are not, and can be handled with relatively simple counseling and stress-management techniques. You can use the following guidelines to help manage the college student stress:

1. understand your role in stress reactions
2. develop a balanced life-style and effective personal organization
3. learn specific relaxation techniques
4. gain perspective on problems by discussing them, and
5. clarify your values and develop a sense of spirituality

There are four primary sources of stress:

The Environment - examples include noise, pollution, traffic and crowding, and the weather.

Physiological - examples include illness, injuries, hormonal fluctuations, and inadequate sleep or nutrition.

Your Thoughts - the way you think affects how you respond. Negative self-talk, catastrophizing, and perfectionism all contribute to increased stress.

Social Stressors - examples include financial problems, work demands, social events, and losing a loved one.

1.2 Data Mining – Introduction:

Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases. Data mining concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. Educational data mining is a new emerging technique of data mining that can be applied on the data related to the field of education. It uses many techniques such as decision trees, neural networks, naive bayes, K-Nearest neighbour and many others. Using these techniques different kinds of knowledge can be discovered using association rules, classification and clustering. By using this we extract knowledge that describes students‟ performance in the end of the semester examination and all their details. In the face of huge amounts of data, the first task is to sort them out, cluster analysis is to classify the raw data in a reasonable way. The so called clustering is a group of physical or abstract objects, according to the degree of similarity between them, divided into

Corresponding Author: R. Priya, Research Scholar, Department of MCA, Vels University, Pallavaram, Chennai, INDIA

several groups, (Srivastava, J., 2000) and makes the same data objects within a groups of high similarity and different groups of data objects which are not similar.

2.The Algorithms in Association Rules Mining:

2.1 The Apriori Algorithm:

The following gives an overview of the Apriori algorithm for finding all frequent itemsets, using the notation. The first pass of the algorithm simply counts item occurrences to determine the large 1-itemsets. A subsequent pass, say pass k, consists of two phases. First, the large itemsets Lk-1 found in the (k-1)th pass are used to generate the candidate itemsets Ck, using the Apriori candidate generation function (apriori-gen) described below. Next, the database is scanned and the support of candidates in Ck is counted. For fast counting, an efficient determination if the candidates in Ck that are contained in a given transaction t is needed. A hash-tree data structure is used for this purpose. The Apriori algorithm is:

Pass 1:

1. Generate the candidate itemsets in C1
2. Save the frequent itemsets in L1

Pass k

Generate the candidate itemsets in Ck from the frequent itemsets in Lk-1

Join Lk-1 p with Lk-1q, as follows: insert into Ck select p.item1, p.item2, . . . , p.itemk-1, q.itemk-1from Lk-1 p, Lk-1q
where p.item1 = q.item1, . . . p.itemk-2 = q.itemk-2, p.itemk-1 < q.itemk-1

Generate all (k-1)-subsets from the candidate itemsets in Ck

Prune all candidate itemsets from Ck where some (k-1)-subset of the candidate itemset is not in the frequent itemset Lk-1

Scan the transaction database to determine the support for each candidate itemset in Ck

Save the frequent itemsets in Lk

The apriori-gen function takes as argument Lk-1, the set of all large (k-1)-itemsets. It returns a superset of the set of all large k-itemsets and is described in.

2.2 The FP-growth Algorithm:

As shown in, the main bottleneck of the Aprioi-like methods is at the candidate set generation and test. This problem was dealt with by introducing a novel, compact data structure, called frequent pattern tree, or FP-tree then based on this structure an FP-tree-based pattern fragment growth method was developed, FP-growth. The definition, according to is as follows.

Definition 1 (FP-tree) A frequent pattern tree is a tree structure defined below.

1. It consists of one root labeled as "root", a set of item prefix sub-trees as the children of the root, and a frequent-item header table.
2. Each node in the item prefix sub-tree consists of three fields: item-name, count, and node-link, where item-name registers which item this node represents, count registers the number of transactions represented by the portion of the path reaching this node, and node-link links to the next node in the FP-tree carrying the same item-name, or null if there is none.
3. Each entry in the frequent-item header table consists of two fields, (1) item-name and (2) head of node-link, which points to the first node in the FP-tree carrying the item-name.

The actual algorithm, according also to (Srivastava, J., 2002) is:

Algorithm 1 (FP-tree construction)

Input: A transactional database DB and a minimum support threshold ξ.

Output: Its frequent pattern tree, FP-tree

Method: The FP-tree is constructed in the following steps:

1. Scan the transaction database DB once. Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items.
2. Create the root of an FP-tree, T, and label it as "root". For each transaction Trans in DB do the following.

a. Select and sort the frequent items in Trans according to the order of L. Let the sorted frequent item list in Trans be [p | P], where p is the first element and P is the remaining list. Call insert_tree([p | P], T).

b. The function insert_tree([p | P], T) is performed as follows. If T has a child N such that N.item-name = p.item-name, then increment N's count by 1; else create a new node N, and let its count be 1, its parent link be linked to T, and its node-link be linked to the nodes with the same item-name via the node-link structure. If P is nonempty, call insert_tree(P, N) recursively.

The FP-growth algorithm for mining frequent patterns with FP-tree by pattern fragment growth is:

Input: a FP-tree constructed with the above mentioned algorithm;

D – transaction database;

s – minimum support threshold.

Output: The complete set of frequent patterns.

Method:

call FP-growth(FP-tree, null).
Procedure FP-growth(Tree, A)
{
if Tree contains a single path P
then for each combination (denoted as B) of the nodes in the path P do
generate pattern $B \cup A$ with support=minimum support of nodes in B

else for each ai in the header of the Tree do
{
generate pattern B = ai ∪ A with support = ai.support;
construct B's conditional pattern base and B's conditional FP-tree
TreeB;
if TreeB ≠ Ø
then call FP-growth(TreeB, B)
}
}

2.3 The DynFP-growth Algorithm:

As shown in (Srivastava, J., 2002)the main bottleneck of the Aprioi-like methods is at the candidate set generation and test. This problem was taken into consideration by introducing a novel, compact data structure, named frequent pattern tree, or FPtree, then based on this structure an FP-tree-based pattern fragment growth method was developed, FP-growth. The completeness and compactness of this structure is also shown in (Srivastava, J., 2002)Some observations on the way the FP-tree are constructed.

1. The resulting FP-tree is not unique for the same "logical" database.
2. The process needs two complete scans of the database.

Algorithm (Dynamic FP-tree construction):

Input: A transactional database DB and a minimum support threshold ξ.

Output: Its frequent pattern tree, FP-tree

Method: The FP-tree is constructed in the following steps:

1. Create the root of an FP-tree, T, and label it as "root". For each transaction Trans in DB the next steps must be followed.

a. Add the items in Trans into the header.

b. Select and sort the items in Trans according to the order of the header's mastertable. Let the sorted frequent item list in Trans be [p | P], where p is the first element and P is the remaining list. Call insert_tree([p | P], T).

c. The function insert_tree([p | P], T) is performed as follows. If T has a child N such that N.item-name = p.item-name, then increment N's count by 1; else create a new node N, and let its count be 1, its parent link be linked to T, and its nodelink be linked to the nodes with the same item-name via the node-link structure. If P is nonempty, call insert_tree(P, N) recursively.

d. If reordering is needed (i.e. a "promotion" was detected) then call reorder() on the FP-tree.

The reorder() function is performed as follows:

1. Gather the "promoted" items into a reorderList ordered according to their support (descending) and lexicographical order.
2. Call checkpoint() to update the insertion order into the FP-tree.
3. For each item from reorderList go through the list of linked nodes and for each of these nodes call moveUp(node) to place that node into the correct position in the FP-tree, according to the header's master-table.

The moveUp(node) function is defined as:

1. Repeat the steps (a. to g.) until the node and its current parent are in the properOrder

a. Take the node's parent's parent (pparent)

b. If parent has the same support as the node, remove the parent from its parent's childNodes and assign it to newNode

c. Else perform the following actions:

i. Create a newNode with the same item as the parent, but having the node's support.

ii. Link it into the parent's list of nodes with the same item.

iii. Adjust the support of the parent, by subtracting the node's support

iv. Remove the node from the childNodes of the parent

d. Replace the childNodes into the newNode with the childNodes from the node and update the parent link of the childNodes with their new parent (newNode).

e. Set the parent link of the node to pparent (the original parent's parent), initialize its childNodes with the newNode, and set the newNode's parent to node.

f. (optional step) If there is already an existingNode for the node's item in the pparent's childNodes, then call merge(existingNode, node), and continue with the existingNode as the current node.

g. Otherwise insert the node into the childNodes of pparent.

The resulting FP-tree is compatible for mining purposes with the original FPgrowth algorithm described in (Srivastava, J., 2002)Because we use the Dynamic-FPtree construction algorithm we renamed the FP-growth in to DynFP-growth.

3. ***Related Work:***

Alaa el-Halees studied how data mining is useful to improve the performance ofthe student in higher education. For this study association rule, classification ruleusing decision tree was used for analysis.

Al-Radaideh *et al*., applied classification data mining techniques to improve thequality of the higher education by evaluating the main attributes of students that affectthe their performance. This study was used to predict the student's final grade in acourse.

Ayesha performed study on student learning behavior. For this factors likeclass quizzes mid and final exam assignment are studied. This study will help thetutors to reduce the ratio of drop out and improve the performance level of students.

Bharadwaj and Pal, used the decision tree method for classification to evaluateperformance of student's. The objective of their study is to discover knowledge thatdescribes students' performance in end semester examination. This study was quiteuseful for identifying the dropout's student in earlier stage and students who needspecial attention and allow the teacher to provide appropriate advising.

Bharadwaj and Pal, conducted study on the student performance based byselecting 300 students from 5 different degree college conducting BCA (Bachelor ofComputer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. Bymeans of Bayesian classification method on 17 attributes, it was found that the factorslike students grade in senior secondary exam, living location, medium of teaching,mother's qualification, students other habit, family annual income and student'sfamily status were highly correlated with the student academic performance.

Boero, Laureti & Naylor, they found that gender (males have a higherprobability of dropping out relative to the reference group of females) is one of theprincipal determinants of the probability of dropping out and age has a significantpositive effect.

Bray, in his study on private tutoring and its implications, observed that thepercentage of students receiving private tutoring in India was relatively higher than inMalaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there wasan enhancement of academic performance with the intensity of private tutoring andthis variation of intensity of private tutoring depends on the collective factor namelysocio-economic conditions.

Cesar *et al.*, proposed a recommendation system based to help students to makedecisions related to their academic track.

Chandra and Nandhini, used the association rule mining analysis to identifiesstudents' failure patterns. The main objective of their study is to identify hiddenrelationship between the failed courses and suggests relevant causes of the failure toimprove the low capacity students' performances.

D'Mello studied on bored and frustrated student.

El-Halees, proposed a case study that used educational data mining to analyzestudents' learning behavior. The objective of his study is to show how useful datamining can be used in higher education to improve student's performance. Theyapplied data mining techniques to discover relevant information from large databasesuch as association rules and classification rules using decision tree, clustering andoutlier analysis.

Fadzilah and Abdullah, applied data mining techniques to enrollment data.Descriptive and predictive approaches were used. Cluster analysis was used to groupthe data into clusters based on their similarities. For predictive analysis, NeuralNetwork, Logistic regression, and the Decision Tree have been used. After evaluatingthese techniques, Neural Networks classifier was found to give the highest results interm of classification accuracy.

Hijazi and Naqvi, conducted as study on the student performance by selecting asample of 300 students (225 males, 75 females) from a group of colleges affiliated toPunjab university of Pakistan. The hypothesis that was stated as "Student's attitudetowards attendance in class, hours spent in study on daily basis after college, studentsfamily income, students mother's age and mother's education are significantly relatedwith student performance" was framed. By means of simple linear regressionanalysis, it was found that the factors like mother's education and student's family.

Oladipupo and oyelade, perform study using association rule data miningtechnique to identify student's failure patterns. They take a total number of 30 coursesfor 100 levels and 200 levels. Their study focuses on constructive recommendation,curriculum structure and modification in order to improve student's academicperformance and trim down failure rate.

Pandey and Pal, conducted study on the student performance based by selecting60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India.By means of association rule they find the interestingness of student in opting classteaching language.

Pathom *et al.*, proposed a classifier algorithm for building Course RegistrationPlanning Model (CRPM) from historical dataset. The algorithm is selected bycomparing the performance of four classifiers include Bayesian Network, C4.5,Decision Forest, and NBTree. The dataset were obtained from student enrollmentsincluding grade point average (GPA) and grades of undergraduate students. As aresult, the NBTree was the best of the four classifiers. NBTree was used to generatethe CRPM, which can be used to predict student class of GPA and consider studentcourse sequences for registration planning.

Ramasubramanian predict aspects of higher education students. In thispaper they analyze that one of the biggest challenges that higher education faces todayis predicting the behavior of students. Institutions would like to know, somethingabout the performances of the students group wise. He proposed a problem toinvestigate the performances of the students when the large data base of Studentsinformation system (SIS) is given. Generally students' problems will be classified intodifferent patterns based on the level of students like normal, average and belowaverage. In this paper we attempt to analyze SIS database using rough set theory topredict the future of students.

Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat Khan applied K-mean clustering to analyze learning behavior of students which will helpthe tutor to improve the performance of students and reduce the dropout ratio to asignificant level.

Shannaq *et al*. applied the classification as data mining technique to predict thenumbers of enrolled students by evaluating academic data from enrolled students tostudy the main attributes that may affect the students' loyalty (number of enrolledstudents). The extracted classification rules are based on the decision tree as aclassification method, the extracted classification rules are studied and evaluatedusing different evaluation methods. It allows the University management to preparenecessary resources for the new enrolled students and indicates at an early stagewhich type of students will potentially be enrolled and what areas to concentrate upon
in higher education systems for support.

Warapon in, who presented the use of data mining techniques, particularlyclassification, to supports high school students in selecting undergraduate programs.Warapon proposed a classification model to give guidelines to students, especially,for the undergraduate programs for making possible better academic plans. Thedecision tree technique was applied to determine which major is best suitable forstudents.

Woodman found for courses in the mathematics and computing faculty at theOpen University in UK, by using the binary logistic regression, that the mostsignificant factors to whether students passed, failed or dropped out, were marks forthe first assignment, the number of math courses passed in the previous two years, thecourse level, the points the course is worth and the occupation group of the student.This was the most parsimonious model, but in the model which includes all 25potential predictors other variables such as ethnicity (ranked as 7th according to itsrelative importance), education (8th), age group (9th), course level (11th), disability(18th) and gender (22nd) were also significant. However, one of the problems withthe logistic regression is that in large samples any difference, may lead to conclusionthat the factor is significant when in fact that is not the case. Using the samemethodological approach with data available at new student registration in the UKOpen University Simpson (2006) found that the most important factor is the courselevel, followed by credit rating of a course, previous education, course programme,socio-economic status, gender and age.

Yadav, Bharadwaj and Pal, To predict the students' performance they obtainedthe university students data like attendance, class test, seminar and assignment marksusing three algorithms ID3, C4.5 and CART.

4. Comparative Studies:

The three frequent pattern mining algorithms were implemented in Java and tested on several data sets. The platform's specifications used for this test was: Pentium 4 3GHz processor, with 1GB RAM, Windows 2000. In order to obtain more realistic results a Microsoft SQL 2000 Server was used and accessed through the standard ODBC interface. To study the performance and scalability of the algorithms generated data sets with 500 to 1000 transactions, and support factors between 5% and 40% were used. Any transaction may contain more than one frequent itemset. The number of items in a transaction may vary, as well as the dimension of a frequent itemset. Also, the number of items in an itemset is variable. Taking into account these considerations, the generated data sets depend on the number of items in a transaction, number of items in a frequent itemset, etc. The necessary parameters to generate the test data sets are defined in Table 1:

Table 1: Parameters to generate the test data sets.

|D|	Number of transactions
|T|	Average size of the transactions
|L|	Number of maximal potentially large itemsets
N	Number of items

The test data set is generated for a number of items N = 100 and a maximum number of frequent itemsets |L| = 1000. |T| was chosen to be 20. Some of the results of the comparison between the Apriori, FP-growth and DynFP-growth algorithms for support factor of 5% and for different data sets are presented in Table 2:

Table 2 shows that the execution time of the algorithms grows with the dimension of the data set. The best performance is obtained by the FP-growth algorithm. Figure 2 shows that the execution time for the FP-growth algorithm is constant for a certain data set when the support factor decreases from 40% to 5% while, in the same time, the execution time of the Apriori algorithm increases dramatically. For a support factor of 30% or greater and a data set of 25,000 transactions, the Apriori algorithm has better performances than the FP-growth algorithm, but for a support factor of 20% or less its performance decreases dramatically. Thus, for a support factor of 5% the execution time for the Apriori algorithm is three times longer than the execution time of the FP-growth algorithm and up to five times longer than DynFP-growth.

The execution time for the two algorithms for different values for the support factor on a data set with 5000 transactions is shown in Table 3. We notice that the Apriori algorithm has a lower performance than the FP-growth and DynFP-growth algorithms even for a support factor of 40%.

Table 2: The results for support factor of 5%.

Transactions (K)	Exectime(sec)		
	Apriori	DynFP-Growth	FP-growth
10	16.97	5.35	5.52
30	24.98	8.65	9.89
50	51.23	11.24	17.55
70	69.5	15.87	23.65
90	108.08	22.51	37.52
110	201.32	39.45	67.84
130	1474.21	55.24	98.25
190	3099.21	99.8	177.84
300	5321.32	153.54	276.52
400	9904.54	287.41	529.88
500	17259.6	461.87	852.45
	20654.4	613.03	1153.24

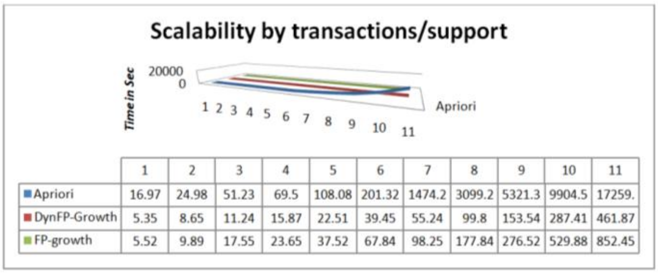

Fig. 1: Scalability by transactions/support.

Table 3: The results for D1 150K by support.

Transactions (K)	Exectime(sec)		
	Apriori	DynFP-Growth	FP-growth
5	3101.23	99.58	173.52
10	2189.03	93.24	169.84
15	1311.58	93.54	169.52
20	1312.25	92.88	167.85
25	875.21	92.89	168.75
30	879.2	92.25	169.374
35	443.25	91.85	166.55
40	444.12	93.89	172.94

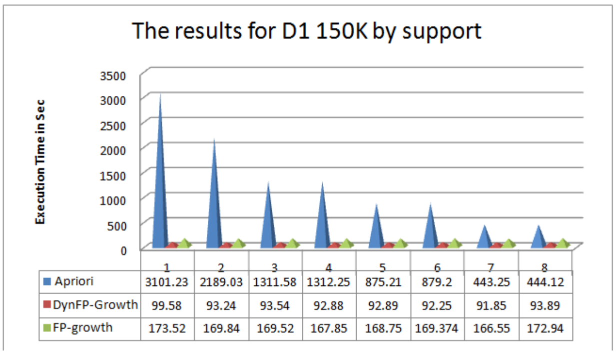

Fig. 2: The results for D1 150 k by support.

Figure 2 presents the execution time of the Apriori algorithm for different values of the support factor on different sized data sets. From here we notice that the performance of the algorithm is influenced by the dimensions of the data set and also by the support factor. Figure 4 presents the execution time of the FP-growth algorithm for different values of the support factor on different sized data sets. From here we notice that the performance of the algorithm is depending only on the dimensions of the data set, the support factor having a very small influence.

Comparison of Prediction Results:

We compared our results of this study with before studies which have done. The comparison results show:

- In an average maximum percentage of students in Indian Engineering colleges belongs to male students.Specially some branches such as Software Engineering and Information Technology in Engineering stream is maximum number of students are female.
-female students had better academic performance in literature and mnemonic science. Whereas, male students had better academic performance in Mathematics and formal science.
- Life style or welfare and parents‟ level of education had various effects on students‟ academic performance. In the other words, female students with low education level of their parent‟s had lower academic performance compare with male students with similar situation.
- Male students were more sensitive than female students when they faced with some unwanted situations and they suffered stress and depression during semester or examination dates more than female students.

6. Conclusion:

Here stressed student datasets analyzed by the above three existing algorithm. From the experimental data presented it can be concluded that the DynFP-growth algorithm behaves better than the FP-growth algorithm. First of all, the FP-growth algorithm needs at most two scans of the database, while the number of database scans for the candidate generation algorithm (Apriori) increases with the dimension of the candidate itemsets. Also, the performance of the FP-growth algorithm is not influenced by the support factor, while the performance of the Apriori algorithm decreases with the support factor. Thus, the candidate generating algorithms (derived from Apriori) behave well only for small databases (max. 5,000 transactions) with a large support factor (at least 32%). In other cases the algorithms without candidate generation DynFP-growth and FP-growth behave much better.

In this investigation, we attempted to discover the effects of some factors such as: the education level of parents, life style or welfare, the number of family members in home, gender and the undertaken branch on students‟ academic performance colleges.

Some additional suggestions for reducing stress levels and enhancing your college experience:

● Keep your space and consequently your mind organized.
● Go to class in time.
● Keep up with course work (the rule of thumb is two hours of study per one hour in class).
● Get involved with campus activities.
● Maintain communication with your family.
● Take advantage of campus resources and choose a career path.
● Form healthy relationships.
● Talk to someone about your problems (family member, friend, college counselor).
● Get to know your professors.

With the help of family, friends, and perhaps campus stress-management resources, many students are able to keep their stress levels relatively under control or even thrive in the college setting.

7. Future Work:

Future work is to study on large database of stressed student in a universityusing other data mining techniques such as Logistic Regression, Clustering andNeural Network in order to determine similarities and relationship betweenmultiple factors.

REFERENCES

Srivastava, J., R. Cooley, M. Deshpande, and Tan, 2000. “Web Usage Mining: Discovery And Applications of Usage Patterns from Web Data‟, SIGKDD, 1(2): 12–23.

Srivastava, J., 2002., “Web Mining: Accomplishments and Future Directions”, University of Minnesota, USA, available at http://www.ieee.org.ar/downloads/Srivastava-tut-paper.pdf (accessed on 23 March 2012).

Stromquist, N.P., 1989. “Determinants of Educational Participation and Achievement of Women in the Third World: A Review of the Evidence and a Theoretical Critique”, Review of Educational Research, 59(2): 143–183.

Oskouei, R.J. and B.D. Chaudhary, 2010. “Internet Usage Pattern by Female Students: A Case Study”,Seventh International Conference on Information Technology: New Generations (ITNG), 12–14: 1247–1250.

Oskouei, R.J., 2010. “Differential Internet Behaviors of Students from Gender Groups”, International Journal of Computer Applications, 4(7): 3638-3642.

Oskouei, R.J., 2010. “ Impact of Non-Academic Websites Usage on Female Students Academic Performance (A Case Study)”, International Conference on Educational and Network Technology (ICENT), 29–33.

Kim, S. and M. Chang, 2007, “ The Differential Effects of Computer Use on Academic Performance of Students from Immigrant and Gender Groups: Implications on Multimedia Enabled Education”, Ninth IEEE International Symposium on Multimedia Workshops, ISMW„07: 479–484.

Koohang, A., 2004. “Students‟ Perceptions Toward the Use of the Digital Library in Weekly Web-Based Distance Learning Assignments Portion of a Hybrid Program”, British Journal of Educational Technology, 35(5): 617–626.

Oskouei, R.J. and Phani Sajja, 2012. “Mining Gender Affinity to Social Networking”, International Journal of Social Network Mining, 1(2): 160-184.

Yao-Guo, G., Lin-Yan, S. and Feng-Lin, 2006. " A Research on Emotion and Personality Characteristics In Junior High School Students With Internet Addiction Disorders", Chinese Journal of Clinical Psychology, 14: 153–155.

Oskouei, R.J. and Phani Sajja, 2012. " Exploring Different Effects of Internet Usages on Female Users" Social and Academic Behaviors" , Journal of Information Technology Review, 2:11-24.

Oskouei, R.J., 2010. " The Role of Social Networks on Female Students Activities", Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, 26-31.

Oskouei, R.J. and B.D. Chaudhary, 2010. " Relationship Between Academic Performance and Usage Pattern of Non-Academic Websites by Gender Group", International Conference on Educational and Network Technology (ICENT), 42-45.

Oskouei, R.J. and Phani Sajja, 2013." Perceived Internet Usage Behaviors as Predictor to Outlier Detection in Students' Communities in Academic Environments", Indian Journal of Science and Technology, 6(7): 4923-4935.

Oskouei, R.J., 2010. " Behavior mining of female students by analyzing log files", Fifth International Conference on Digital Information Management (ICDIM), 5–10.

http://computation.llnl.gov/casc/sapphire/overview/overview.html

Baradwaj, K., saurabh pal, 2011. "Mining Educational Data to Analyze Students" Performance ", International Journal of Advanced Computer Science and Applications (IJACSA), 2-6.

Chai, K., H.T. Hn, H.L. Chieu, 2002. "Bayesian Online Classifiers for Text Classification and Filtering", Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, 97-104.

Jain, A.K., M.N. Murty, P.J. Flynn, 1999. "Data Clustering: A Review", ACM Computing Surveys, 31-3.

Springer Data mining Book, chapter 8, "Cluster Analysis: Basic Concepts and Algorithms ", 487-569.

U.S. Headquarters: StatSoft, 2007. "New Features In Statictica 8",

http://www.statsoft.com/Portals/0/pdf/statistica8f eatures.pdf. (Accessed 31 March 2014)

Donald, K., Wedding, 2014. "Extending the Data Mining Software Packages SAS Enterprise Miner and SPSS Clementine to Handle Fuzzy Cluster Membership: Implementation with Examples"ÂÌ,http://web.ccsu.edu/datamining/DataMining/Theses/Don/thesis.pdf. (Accessed on 31 March 2014)

http://www.support.sas.com

Minaei-Bidgoli, B., D.A. Kashy, G. Kortemeyer and W.F. Punch, 2003. " Predicting Student Performance: An Application of Data Mining Methods With An Educational Web- Based System.", In Proceedings of The 33rd ASEE/IEEE Frontiers in Education Conference,(T2A), pp. 13-18, doi: 10.1109/FIE.2003.1263284.

Machado, L.D.S. and K. Becker, 2003. " Distance Education : A Web Usage Mining Case Study for the Evaluation of Learning Sites." , In Proceedings of The 3rd IEEE International Conference on Advanced Learning Technologies ,(ICALT03), 360-361.

Merceron, A. and K. Yacef, 2005. "TADA-Ed for Educational Data Mining." Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 7-1.

Leiba, M. and R. Nachmias, 2006. " Web Usage Patterns and Learning Styles in An Academic Course In Engineering.", International Conference on Information Communication Technologies in Education. (ICICTE).

Liccardi, Ilaria, Ounnas, Asma, Pau, Reena, Massey, Elizabeth, Kinnumen, Paivi, Lewthwaite, Sarah, Midy, Marie-Anne and Sakar, Chandan, 2007. " The Role of Social Networks in Students Learning Experiences." , In Proceeding of the Working group reports on Innovation and technology in computer science education, (ITiCSE-WGR "07), ACM New York, NY, USA, pp.224-237.

Figl, K., S. Kabicher, K. Toi, 2008. "Promoting Social Networks Among Computer Science Students. " , In Proceedings of the 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Springer, 15-20.

Welty, K., 1996. "Identifying Women"s Perspectives on Technology" Paper presented at the International Technology Education Association Conference in Phoenix, AZ.

Jo Sanders, 2005. " Gender and Technology in Education: A Research Review", In Proceedings of The international symposium on Women and ICT: creating global transformation (CWIT "05), ACM.

Kim, S. and M. Chang, 2007. " The dierential Eorts of Computer Use on Academic Performance of Students from Immigrant and Gender Groups: Implications on Multimedia Enabled Education. " , In Proceeding of the Ninth IEEE International Symposium on Multimedia Workshops IEEE Computer Society Washington, (ISMW "07).

Qasem, A,. A-Radaideh, E. A-Shawakfa and M. AI-Najjar, 2006. "Mining Students Data Using Decision Tress.", (2006), International Arab Conference on Information Technology , (ACIT2006), 1-5.

Jamili Oskouei, R., 2010. "Analyzing Dierent Aspects of Social Network Usages on Students Behaviors and Academic Performance" , IEEE, International Conference on Technology for Education (T4E), 216-221.

Alaa El-Halees, 2008. "Mining Students Data to Analyze Learning Behavior: A Case Study", https://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf (Accessed on

Gang Wang, 2014. “A Novel Data Mining Algorithm for Mathematics Teaching Evaluation”, TELKOMNIKA Indonesian Journal of Electrical Engineering, 12(4): 2658-2666.

Xiaodong Zhu, 2014. “On Data Mining Technology to the Quantitative Efficiency Assessment using SBM Model: An Empirical Study on Education Efficiency in Jiangxi Province”, 12(3): 1933-1938.

Braten and H. Stromso, 2006. "Epistemological Beliefs, Interest and Gender as Predictors of Internet-Based Learning Activities", Education Miscellaneous Papers Computers in Human Behavior, 22(6): 1027-1042.

Zhou, B., S.C. Hui and A.C.M. Fong, 2005. “Discovering and Visualizing Temporal-based Web Access Behavior. ", In Proceedings of The IEEE/WIC/ACM International Conference on Web Intelligence, 297-300.

Nilufer Ozabac, 2010. “The effectiveness of teachers to recognize the symptoms of depression for their depressivestudents”,doi:10.1016/j.sbspro.2010.03.339,2010, 2371–2376.