Extraction of Dominant Attributes and Guidance Rules for Scholastic Achievement Using Rough Set Theory in Data Mining

P.G. JansiRani[1] and R. Bhaskaran[2]

**[1]Department of Mathematics, Sethu Institute of Technology
Virudhunagar-626115, Tamil Nadu, India.**

**[2]School of Mathematics, Madurai Kamaraj University
Madurai-625021, Tamil Nadu, India.**

Abstract

Extracting hidden information from a huge set of data is an important and a challenging task in data mining. Data with credibility and relevance plays a vital role in this task. Not to get sidetracked, it is important to ensure that genuine and good quality data is being used. This paper brings out the significance and highlights the efforts for the collection of genuine and good quality data of the academic performance of students by carrying out many pilot and through studies. with different types of questionnaires. An algorithm, using rough set theory, is developed to implement in the data collected from various college students to identify the dominant attributes to be strengthened. Reducts and guidance rules, using rough set theory in data mining, are computed to facilitate the educators and high and low academic achievers to identify the attributes to be strengthened to have scholastic achievement.

Keywords: *Data mining, rough set theory, indiscernity, dominant attributes, scholastic achievement.*

1. Introduction

Everybody makes mistakes. Trying to hide our shortcomings will not make them go away. But an intelligent person learns from past which leads to growth and better capabilities. It has been revealed [10][3]that type of school type, educational level of parents, consumption of alcohol, smoking, time reserved for homework, consumption of vegetable parental head circumference(HC) and parental nutritional indicators are the most important independent variables that determine children's HC and disorders, related to brain volume and IQ.

Method, based on the application of optimization heuristic and visualization of the multidimensional data, evaluates the level of mathematization of different computer science subjects or their nearness to the humanities[4].In another study [12], based on rough set theory, reveals that good performance in pre-university tests implies good performance in university examinations. Also, the number of hours spent for mathematics, sciences and literature in the last year of secondary education and factors like caring, dedicated teachers and a positive atmosphere and support services have been extracted as the factors responsible for the achievement of higher performance in schools [9], while prediction of the exact attributes and the formula for student's success which enhance or reduce the performance of a student is still a challenge. Data mining is discovery driven, a hypothesis (rule) or model, revealing dependencies are automatically extracted from data between condition and decision variables, which will be helpful to make decisions faster also give guidance rules to have better performance in future, in medical diagnosis and in other fields. Eight significant attributes - admission exam score, personal responsibility, first grade obtained at the faculty, good time management and ability to allocate time for studying, learning style and preparation for classes and in-class activity have been identified to discriminate between the more successful students and the less successful[6].

2. Basic Concepts of Rough set theory

2.1 Information System

An information system is defined [7] as a set of objects $S = \{ U, A, V, f \}$ where in this study, U : set of students(objects), A : Various parameter influencing their academic performance(attributes), V : The value given to the parameter (yes (or) No or 5 Point scale)defined by the relation f as $f : U \times A \rightarrow V$

2.2 Discernibility and Indiscernibility

Discernibility and Indiscernibility [7] relation of objects in the information system in rough set theory are two main concepts which are very effective in classification, characterization and clustering the objects. Discernibility matrix consist of entries or set of attributes which discern two objects, while the indiscerniblity matrix IDxy consists entries of attributes which are common to two objects defined by
IDxy $= \{a \in A : f(a,x) = f(a,y)\}$ where $x,y \in U \}$

2.3 Lower and Upper bounds

Let $X \subseteq U$ set of objects, $P \subseteq A$ set of attributes, P – lower approximation[7] of X contains all objects that with the knowledge of attributes P can be classified as certainly belonging to concept X.P – upper approximation of X contains all objects that with the knowledge of attributes P cannot be classified as not belonging to concept X. This will enable us to identify the attributes to derive association rules, relating attributes and decision variables.

2.4 Reduct

Reduct is the minimal set of attributes preserving classification power on original data set A[8] Intersection of all reducts is called core.

3. Methodology

Methodology on data Collection

Questionnaires can be used to collect data quite quickly, directly and openly. But no survey can achieve success without a well-designed questionnaire. Unfortunately, questionnaire design has no theoretical base to guide researcher in developing a flawless questionnaire. Researcher can guide only with the past and present experience of themselves and other researchers. Hence, questionnaire design is more of an art than a science.

Therefore in this study significance efforts have been taken by carrying out many pilot studies with different types of questionnaires prepared before finalizing the final questionnaire for the collection of genuine data on the academic performance of academically high and low achievers.

Questionnaire Model 1

Data are collected through Questionnaire- Model 1 on various activities of students from 5 A.M to 1 A.M.

Questionnaire Model 2

Data are collected through the Questionnaire- Model 2 with 68 questions with maximum of 5 options in each question (either in terms of % or in terms of linguistics), from students who are divided into two clusters where cluster 1 represents students who are good in academic achievement and cluster 2 represents students who are low in academic performance.

From the results, extracted using indiscernity matrix, it is observed that the combination of attributes having majority voting for the low and high achievers are the same, and some of the dominant attributes selected by the low achievers are contradictory, resulting in an inconclusive result[11] and doesn't lead to extract most probable attributes responsible for high and low performance. Therefore we modified the questionnaire model 1 and 2 into model 3.

Questionnaire Model 3

Generally, feedback is got from students in the printed format with 20 questions with grades in the middle and at the end of the semester to assess the quality of teaching rendered and received by the faculty and the students respectively. But students did not take this method very serious and result is not reliable. Therefore it was tried in a plain paper without identity and without any pre printed questions. This resulted in a successful collection of reliable and honest information. This encouraged to collect data responsible for high or low academic achievement through a questionnaire with open ended questions and with no answers to circle, which will not force the respondents to select an answer.

Questionnaire Model 4

Since students showed genuine interest in answering the questions in the modified questionnaire-model-3, we further modified it and framed the Questionnaire Model 4 .In Model 4 questionnaire, the feedback form was used to write descriptively, in the following 3 categories without mentioning their identity which will facilitate to answer frankly and honestly:
a. Self contribution/responsibility for their high achievement/ low achievement.
b. Parental contribution/responsibility for their high achievement / low achievement.
c. Society, institute contribution/responsibility for high achievement/ low achievement.

We got incredible response from students in this model.In this study those who scored more than 1000 marks in Higher Secondary course were considered as toppers since getting 1000 marks is considered, a good achievement in Tamil Nadu, India which is also the prime factor to decide their career in future. Previous study [30] found that the number of hours of mathematics, sciences and literature in the last year of secondary education are more important to estimate the chances of success.

Then these descriptive raw data were transformed into an exhaustive collection of around 300-400 attributes. From the exhaustive collection of raw attributes, closely related attributes were grouped, given in Table 1 and Table2 and each group is labeled with an identity

We found a major difference, among the students while filling up the questionnaires of model 1 and 2 and model 3 and 4. For model 1 and 2 they showed

less interest and less genuinity, and took less, so reliable and definite reasons could not be inferred from model 1 and 2 .

While filling up the questionnaire model 3 and 4 response was really appreciable. Low achievers took it as a chance to express their grievances to make

Table 1. High achievers grouped attributes

A1	S1,S20,	Awareness, poverty Goal passion for studies Previous failure –motivation self interest, self motivation to join in a good institutes to get 1st grade, willpower, poverty
A2	S2,S24	Senior's guidance, Friends, good roommates, helping friends, competition between friends, co-ordination of friends, group study, by teaching to others
A3	S3,S6,S8,S10,S28, S33	More revision, Hard work, dedication, sincere before exam, sincere, fear about exam , placements, study for class test, study thoroughly, careful, frequent test, revision, exam fear, Study in the last time.
A22	S4	Tuition
A4	S5,S13	God, Prayer
A5	S7,S26	Hobbies, Sacrifice, TV, No mobiles
A6	S9,S19,S34, S11	Understanding, concentration in basic concepts, Listening, taking notes, previous year question paper, more reference books, circular books, exam oriented studies understanding concepts, assignment day to day incidence, concepts(Basic)practical knowledge, guidance.
A7	S12,S32, S35	Confidence during exam, happy mind, no tension, no exam fear, not studying new portion on exam day, optimist, don't compare with others, good thoughts, happy mind, no jealous, relaxation.
A23	S14	Presentation.
A24	S15	Regular studies, planning, sincere.
A8	S17,S25	Understanding good relationship between teachers and students, good environment.
A25	S22	Good memory.
A26	S27	Respect, obedience.
A27	S29	Extra studies in early morning & Late night, less sleep, utilizing study holidays.
A28	S30	Regular attendance.
A29	S31	More concentration for problem papers, zeal for state 1st rank, writing practice.
A9	P1, P5, P12, P14, P23	Affection, care during exams, constant care, encouragement expectation, setting goal in mind, guidance, motivation, create positive mentality, parent - student relationship, prizes, proper guidance encouragement @ failure.
		Correcting mistakes, no comparison, not so strict, expectations, confidence, advise, frequent tips, make to realize future financial conditions, difficulties in family situations, parents not educated.
A10	P2, P9, P6, P16, P17	Friendly, co-operation while studying, mother-doubt clarification, good relationship, mother-teaching, freedom, guidance, no pressure.
		Freedom, free in other things, insisting not to play while studying, strict in studies
A11	P3,P10,P19	Present whenever needed, no house work, moral support during morning and night study, any help, left all their desires sacrifice TV, sleeping time, fathers sacrifice for tuition.
A30	P4	care for food, health care
A12 A13	P7,P21, P8, P18	Good environment, no tension, Parents stopped quarreling, worries, provide relaxation facilities, good school, tuition, allowed to stay in hostel, all books, constant support in sufferings, financial support, help during exams, previous year question papers.
A31	P11	All efforts, special care, care and concentration, conducting frequent enquiries, insisting moral values parents -teachers interaction.
A32	P13	Roll models
A33	P15	Prayer , blessings
A34	P22	Trust in children
A14	P24,P19	Tuition, no housework.
A15	T1,T16	Clearing doubts, excellent lectures, good teaching, solving problems easily, make to understand basic concepts, doubt clarification.
A35	T2	Frequent tests, tough tests.
A16	T3, T17, T24, T7, T8 , T28	Advise, correcting mistakes, encouragement, guidance, motivation, teaching in examination point of view, placement, prizes, special care, tips, financial support, motivation for low marks.
		Prayer, confidence, made to realize responsibilities and values.
A17	T4, T13, T9, T15	Morning and evening study, notes, special classes, special coaching, dedication, devoted hard work by staff, punctual, sincere, advice, care.
		Helping tendency, involvement, teachers interest, teachers sacrifice, coaching during holidays, hard work.
A18	T5,T21,T23	Co-operation & care, good relationship between teachers and students, good environment, no partiality.
A19	T6,T19	Affection, discipline, valuation, strict, roll model.
-----	-----------	--
A39	T22, T23	no partiality, strict

Table 2. Low achievers grouped attributes

B38	LS1	Poor handwriting, Language Problem, Poor Presentation, Presentation, (Tamil) Medium problem.
B1	LS2,LS4,LS13	Not studying in exam days also, no proper learning, no syllabus coverage, Not serious in Studies, not practiced problematic subjects, No recap, No Study, No revision, no writing practice for problem oriented subjects.
B14	LS3	No proper planning, no preparation, unable to cover full syllabus.
B2	LS5,LS11	Tiredness, no seriousness, laziness, carelessness, lethargic, more time spent on sleeping, not listening in the class, less hard work, no hard work.
B3	LS6,LS19	More time spent for Sports Activities, Distraction in thoughts, Diversion, games, playing, extracurricular activities, Different Attitude.
B15	LS7	No concentration, more travelling to other places, poor understanding, no concentration, mug up, Less concentration.
B16	LS8	Over confidence, no confidence.
B17	LS9	Less memory.
B4	LS10,LS12, LS30	Studying only during exam time, last minute study, studying during previous day of exam, no daily study, no regular study, postponing of subjects, over load.
B5	LS14,LS18	Roaming, Going out with friends, chatting, cell phones, discussing with friends wasting time, enjoying with friends, meeting friends, no serious effort, TV, wasting time.
B18	LS15	Tension during exam time, fear.
B19	LS16	No meditation.
B6	LS17,LS20,LS25,LS29	Stress, Health Problem, DEPRESSION, FINANCE PROBLEM, Personal problem.
B20	LS21	Lack of attendance, Not attending the classes.
B21	LS22	No interest, No perfect Goal
B22	LS23	NO NOTES, not referring books, lack of advice
B7	LS24,LS26, LS28	Home sick, new environment, new school, new teachers.
B23	LS27	No good friends
B24	LP1	No Extra Effort.
B25	LP2	Economic problem, financial problem.
B26	LP3	Very strict.
B27	LP4	Personal problem.
B8	LP5,LP6	Disturbances during studies, Bad Environment, More guest at home.
B9	LP7,LP9, LP11,LP12	Allow to watch TV, More freedom, not strict, no care, less care, carelessness, disappointment, hostel, no affection, no early morning wake up, not sacrificing excess TV time, Roaming.
B28	LP10	Comparing with others, no motivation, no encouragement.
B29	LP14	No trust in the student
B30	LP15	Uneducated parents.
B31	LT1	Less study holidays.
B10	LT2,LT3	No guidance, no motivation, no encouragement by staff, no co-operation.
B13	LT4,LT10,LT13	Language problem, no good relationship, no interaction between staff and hesitation.
B32	LT7	Friendly relationship.
B12	LT8,LT17	Over strict, punishment, over pressure.
B33	LT9	More homework, no home work, more working hrs.
B34	LT12	No lab facility, no good infrastructure
B35	LT16	Bad timings
B36	LT19	No frequent test conducted
B37	LT20	Partiality in marks

Source of Data: Alagappa Chettiar College of Engineering and Technology Tamil Nadu, India,, Sethu Institute of Technology, Tamil Nadu, India., Thiagarajar College of Engineering, Tamil Nadu, India. Yadava Women's Arts and Science College, Tamil Nadu, India.

self analysis and attribute their mistakes for their lower performance. Top achievers filled it with great interest also expressed gratitude to their parents, teachers and friends who were behind their success which helped us very much in mining a reliable reasons for the success and failure of students.

3.4 Algorithm to Compute Indiscernity matrix and dominant attributes

The algorithm is developed to compute the indiscernity matrix and dominant attributes using PHP language.

Step1: Data is entered in CSV format.

Step2: The set of options for all questions selected by a student is considered as array 1 and that of the next student is considered as array 2.

Step3: Indiscernity matrix has computed from the data matrix by using row intersect Mysql query.

Step4: These indiscern values are inserted twice into column S1 of the data base table, and the corresponding row and column identities are entered in column S2.

Step5: Distinct entries of S1 and S2 are selected from the table sort.

Step6: The count and percentage corresponding to each entry is computed.

Step7: The combination of attributes above the required % are selected as a dominant attributes from

cluster1 (good achievers) students to be strengthened by cluster2 (low achievers) students.

3.5 Dominant attributes extracted from achievers data

Table 3 gives dominant attributes which are extracted from the data of achievers using algorithm specified in section3.4

3.6 Anticipatory warning attribute extracted from low achievers data

Same procedure is applied on the data of low achievers to extract warning attributes given in Table4 to be strengthened in to improve their performance.

Table 3. Clustered dominant attributes high achieves

Attributes	No. of Votes	% of Voting	Name of the Attributes
A91, A101	348	74.0426	Affection, care during exams, constant care, encouragement expectation, future goal, guidance, motivation, create positive mentality, parent - student relationship, prizes, proper guidance, encouragement @ failure. correcting mistakes, no comparison, not so strict, expectations, confidence, advise, frequent tests, make to realize future financial conditions, difficulties in family situations, parents not educated ,friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure.
A91	327	69.5745	Friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure.
A31	251	53.4043	more revision, Hard work, dedication, sincere before exam, sincere, fear about exam , placements, study for class test, study thoroughly, careful, frequent test, revision, exam fear, Study in last time
A151	238	50.6383	Clearing doubts, excellent lectures, good teaching, solving problems easily, make to understand basic concepts, basic concepts, doubt clarification.
A101	229	48.7234	friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure
A91, A151	206	43.8298	Friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure.Clearing doubts, excellent lectures, good teaching, solving problems easily, make to understand basic concepts, basic concepts, doubt clarification.
A91, A101, A151	203	43.1915	Friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure.Clearing doubts, excellent lectures, good teaching, solving problems easily, make to understand basic concepts, basic concepts, doubt clarification friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure
A171	198	42.1277	Morning and evening study, notes, special classes, special coaching, dedication, devoted hard work by staff, punctual, sincere, advice, care.
A31, A91, A101	195	41.4894	More revision, Hard work, dedication, sincere before exam, sincere, fear about exam , placements, study for class test, study thoroughly, careful, frequent test, revision, exam fear, Study in last time affection, care during exams, constant care, encouragement expectation, future goal, guidance, motivation, create positive mentality, parent - student relationship, prizes, proper guidance, encouragement @ failure. correcting mistakes, no comparison, not so strict, expectations, confidence, advise, frequent tips, make to realize future financial conditions, difficulties in family situations, parents not educated ,friendly, co-operation while studying, mother doubt clarification, good relationship, mother teaching, freedom, guidance, no pressure.

Table 4. Clustered dominant attributes low achieves

Attributes	No. of Votes	% of Voting	Name of Attributes
b341	153	70.5069	No good infrastructure
b371	139	64.0553	Partiality in marks
b361	98	45.1613	No frequent test conducted
b21	96	44.2396	Tiredness, no seriousness, laziness, carelessness, lethargic, no sleep, not listening in the class, less hard work, no hard work.
b351	91	41.9355	Bad timings
b21,b371	80	36.8664	Tiredness, no seriousness, laziness, carelessness, lethargic, no sleep, not listening in the class, less hard work, no hard work. partiality in marks
b111	78	35.9447	No syllabus coverage, no clear explanation, capacity is less, fast teaching, no doubt clarification, no good coaching, no good college, no teacher, no proper teaching, not well versed, poor teaching, no proper notes, no special care, no senior staff, critical syllabus, tough syllabus, no special coaching, no practical teaching, no regular attendance, no special assignments, wave length mismatch.
b21,b341	78	35.9447	Tiredness, no seriousness, laziness, carelessness, lethargic, no sleep, not listening in the class, less hard work, no hard work. no lab facility, no good infra
b21,b351	70	32.2581	Tiredness, no seriousness, laziness, carelessness, lethargic, no sleep, not listening in the class, less hard work, no hard work. bad timings
b51	67	30.8756	Roaming, going out with friends, chatting, cell phones, discussing with friends wasting time, enjoying with friends, meeting friends, no effort, TV, wasting time.
b91	51	23.5023	Allow to watch TV, more freedom, not strict, no care , less care, carelessness, disappointed, hostel, no affection, no early morning wake up, not spending time,no TV sacrifice, going outside.

4. Follow-up study on the attributes influencing academic achievement after four years.

Based on the dominant attributes extracted from the study done in the year 2006 using questionnaire model 4, Questionnaire model 5 based on the attributes given in table 6, is prepared and data were collected in the year 2010 from 566 students of

Table 5. Attributes of Model 5 Questionnaire

Attribute Number	Name of Attribute
1	Parents Motivation
2	Self ambition and enthusiasm
3	Laziness and postponing studies
4	Study environment at home
5	Regularly studying habit
6	Teachers motivation and efforts
7	Concentration the class
8	Memory capacity
9	Time spent on Watching TV, Chatting with friends, Roaming
10	Health problem affecting academic progress
11	Family financial problem affecting academic progress
12	School timings affecting academic progress
13	Drinking water per day
14	Revising more number of times
15	Last minute study only

various schools located in various places in Tamilnadu. The data were analyzed and reducts and generation rules [2] were extracted using the data mining software Rough Set Exploration System (RSES). It is very significant and interesting to note, from the comparative study on the dominant attributes extracted in the academic data collected in the year 2006 and in the year 2010, that factors influencing academic achievement doesn't change much over time. Previous studies [11] using weka data mining soft ware revealed that gender, interest for mathematics and performance in English are the dominant predicting attributes while other search algorithms predicts only mathematics and performance in English as important predictors. Besides dominant attributes extracted from the study done in the year 2006 as a follow-up study reducts, guidance rules and anticipatory warning rules have been extracted from the academic data collected in the year 2010 to predict [1] the expected performance of students using RSES software.

4.1 Computation of reducts and generation of guidance rules

Reducts computed through Exhaustive and Genetic algorithm

S. No.	Size	Pos. Reg	SC	Reducts
1	13	0.951	1	{1,2,3,4,5,6,7,8,9,10,13,14,15}
2	13	0.951	1	{2,3,4,5,6,7,8,9,10,11,13,14,15}
3	13	0.951	1	{2,3,4,5,6,7,8,9,10,12,13,14,15}

Rules having minimum support of 3 using covering algorithm

(1-10)	Match	Decision rules
1	10	(3=1)&(9=0)&(8=1)&(7=1)=>(mark={A[10]})
2	9	(5=1)&(14=1)&(3=1)&(4=1)&(13=1)&(9=1)&(8=1)&(7=1)=>(mark={B[4],...
3	6	(3=0)&(9=0.5)&(8=1)&(7=1)=>(mark={B[6]})
4	6	(3=0.5)&(9=0.5)&(8=0.5)&(7=0.5)=>(mark={A[6]})
5	5	(3=0)&(13=1)&(9=1)&(8=1)&(7=1)=>(mark={A[5]})
6	5	(3=0.5)&(13=1)&(9=0)&(8=1)&(7=0.5)=>(mark={A[5]})
7	4	(3=0)&(13=1)&(9=0)&(8=1)&(7=0.5)=>(mark={B[4]})
8	4	(3=0)&(9=0.5)&(8=0.5)&(7=1)=>(mark={A[4]})
9	4	(3=0)&(9=1)&(7=0.5)=>(mark={A[4]})
10	3	(3=1)&(9=0)&(8=0)&(7=1)=>(mark={A[3]})

Rules having minimum support of 13 using exhaustive algorithm

(1-114)	Match	Decision rules
1	25	(4=1)&(7=1)&(8=1)&(9=0)&(10=1)=>(mark={A[25]})
2	24	(4=1)&(5=1)&(7=1)&(9=0)&(11=1)=>(mark={A[24]})
3	23	(5=1)&(6=1)&(7=1)&(9=0)&(11=1)=>(mark={A[23]})
4	22	(1=1)&(2=1)&(7=1)&(8=1)&(9=0)&(10=1)&(11=1)=>(mark={A[22]})
5	22	(2=1)&(7=1)&(8=1)&(9=0)&(10=1)&(11=1)&(12=1)=>(mark={A[22]})
6	20	(7=1)&(9=0)&(10=1)&(14=1)=>(mark={A[20]})
7	20	(1=1)&(2=1)&(3=0.5)&(6=1)&(8=1)&(9=0)&(11=1)=>(mark={A[20]})
8	19	(2=1)&(3=0.5)&(6=1)&(8=1)&(9=0)&(11=1)&(12=1)=>(mark={A[19]})
9	19	(5=1)&(7=1)&(8=1)&(9=0)=>(mark={A[19]})
10	18	(2=1)&(4=1)&(7=1)&(9=0)&(10=1)&(13=1)=>(mark={A[18]})
11	18	(2=1)&(6=1)&(7=1)&(9=0)&(10=1)&(13=1)=>(mark={A[18]})
12	17	(2=1)&(4=1)&(7=1)&(12=0)=>(mark={A[17]})
13	17	(4=1)&(5=1)&(7=1)&(9=0)&(10=1)=>(mark={A[17]})
14	17	(5=1)&(6=1)&(7=1)&(9=0)&(10=1)=>(mark={A[17]})
15	17	(6=1)&(7=1)&(9=0)&(10=1)&(15=1)=>(mark={A[17]})
16	16	(3=0.5)&(4=1)&(8=0.5)&(11=1)&(12=1)&(14=1)=>(mark={A[16]})
17	16	(2=1)&(3=0.5)&(5=0)&(8=1)=>(mark={A[16]})
18	16	(3=0.5)&(4=1)&(8=0.5)&(10=1)&(14=1)=>(mark={A[16]})
19	16	(2=1)&(3=0.5)&(4=1)&(5=0)&(10=1)=>(mark={A[16]})
20	16	(1=1)&(7=1)&(8=1)&(9=0)&(14=0)=>(mark={A[16]})
21	16	(7=1)&(8=1)&(9=0)&(12=1)&(14=0)=>(mark={A[16]})
22	16	(2=1)&(3=0.5)&(4=1)&(6=1)&(8=1)&(9=0)&(11=1)=>(mark={A[16]})
23	16	(5=1)&(6=1)&(7=1)&(9=0)&(13=1)=>(mark={A[16]})
24	16	(5=1)&(7=1)&(9=0)&(11=1)&(14=1)=>(mark={A[16]})
25	16	(4=1)&(5=1)&(7=1)&(9=0)&(13=1)=>(mark={A[16]})
26	16	(4=1)&(7=1)&(9=0)&(10=1)&(15=1)=>(mark={A[16]})
27	15	(4=1)&(14=0.5)&(15=1)=>(mark={A[15]})
28	15	(2=1)&(5=0.5)&(7=0.5)&(9=0.5)=>(mark={A[15]})
29	15	(1=1)&(5=0.5)&(9=0.5)&(10=1)&(12=1)=>(mark={A[15]})
30	15	(7=0.5)&(9=0.5)&(11=1)&(15=1)=>(mark={A[15]})
31	15	(2=1)&(7=0.5)&(9=0.5)&(15=1)=>(mark={A[15]})
32	15	(3=0.5)&(4=1)&(8=0.5)&(12=1)&(13=1)&(14=1)=>(mark={A[15]})
33	15	(2=1)&(3=0.5)&(5=0)&(6=1)&(10=1)&(13=1)=>(mark={A[15]})
34	15	(2=1)&(3=0.5)&(5=0)&(6=1)&(12=1)&(13=1)=>(mark={A[15]})
35	15	(1=1)&(2=1)&(3=0.5)&(7=1)&(9=0)&(10=1)=>(mark={A[15]})
36	15	(1=1)&(2=1)&(3=0.5)&(7=1)&(8=1)&(9=0)=>(mark={A[15]})
37	15	(2=1)&(3=0.5)&(7=1)&(9=0)&(10=1)&(12=1)=>(mark={A[15]})
38	15	(4=1)&(5=0)&(6=1)&(10=1)&(13=1)&(14=1)=>(mark={A[15]})
39	14	(2=1)&(5=0)&(8=1)&(9=1)=>(mark={A[14]})
40	14	(2=1)&(4=1)&(7=1)&(14=0.5)=>(mark={A[14]})

Rules having minimum support of 13 using exhaustive algorithm Contd.

41	14	(7=0.5)&(9=0.5)&(11=1)&(14=1)=>(mark={A[14]})
42	14	(2=1)&(5=0.5)&(9=0.5)&(10=1)&(12=1)=>(mark={A[14]})
43	14	(3=0.5)&(6=1)&(7=0.5)&(9=0.5)&(11=1)=>(mark={A[14]})
44	14	(2=1)&(5=0.5)&(9=0.5)&(15=1)=>(mark={A[14]})
45	14	(2=1)&(3=0.5)&(4=1)&(5=0)&(6=1)&(12=1)=>(mark={A[14]})
46	14	(2=1)&(3=0.5)&(7=1)&(8=1)&(9=0)&(12=1)=>(mark={A[14]})
47	14	(2=1)&(3=0.5)&(4=1)&(7=1)&(8=1)&(9=0)=>(mark={A[14]})
48	14	(2=1)&(3=0.5)&(4=1)&(7=1)&(9=0)&(10=1)=>(mark={A[14]})
49	14	(3=0.5)&(4=1)&(7=1)&(8=1)&(9=0)&(12=1)=>(mark={A[14]})
50	14	(3=0.5)&(5=0)&(6=1)&(10=1)&(14=1)=>(mark={A[14]})
51	13	(2=1)&(3=0.5)&(6=1)&(14=1)&(15=0)=>(mark={A[13]})
52	13	(4=1)&(7=1)&(9=1)&(10=0)=>(mark={A[13]})
53	13	(4=1)&(5=1)&(7=1)&(10=0)&(11=1)=>(mark={A[13]})
54	13	(1=1)&(5=0.5)&(7=0.5)&(9=0.5)=>(mark={A[13]})
55	13	(2=1)&(6=1)&(7=0.5)&(9=0.5)&(13=1)=>(mark={A[13]})
56	13	(6=1)&(7=0.5)&(9=0.5)&(11=1)&(13=1)=>(mark={A[13]})
57	13	(3=0)&(7=1)&(9=0)&(10=1)&(13=1)=>(mark={A[13]})
58	13	(3=1)&(6=1)&(7=1)&(9=0)=>(mark={A[13]})
59	13	(5=0.5)&(9=0.5)&(10=1)&(15=1)=>(mark={A[13]})
60	13	(2=1)&(3=0.5)&(5=0)&(6=1)&(9=0)=>(mark={A[13]})
61	13	(2=1)&(3=0.5)&(7=1)&(8=1)&(10=0)=>(mark={A[13]})
62	13	(2=1)&(3=0.5)&(4=1)&(5=0)&(12=1)&(13=1)=>(mark={A[13]
63	13	(3=0.5)&(4=1)&(5=0)&(10=1)&(14=1)=>(mark={A[13]})
64	13	(6=1)&(10=1)&(14=0.5)&(15=1)=>(mark={A[13]})
65	13	(8=1)&(14=0.5)&(15=1)=>(mark={A[13]})
66	13	(2=1)&(7=1)&(8=1)&(9=0)&(10=1)&(13=1)=>(mark={A[13]})
67	13	(2=1)&(3=0.5)&(5=0)&(6=1)&(12=1)&(14=1)=>(mark={A[13]

Rules having minimum support of 3 using LEMS algorithm

(1-204)	Match	Decision rules
1	13	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(9=0)&(3=0.5)&(7=1)=>(mark={A[13]})
2	13	(1=1)&(11=1)&(2=1)&(4=1)&(6=1)&(15=1)&(7=1)&(12=0)=>(mark={A[13]})
3	12	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(14=1)&(3=0.5)&(8=0.5)=>(mark={A[12]})
4	10	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(8=1)&(7=1)&(15=1)&(9=0)=>(mark={A[10]})
5	9	(1=1)&(11=1)&(12=1)&(4=1)&(6=1)&(10=1)&(2=1)&(13=1)&(9=1)&(7=1)&(15=1)&(14=1)&(3=1)&(5=1)&(8=1)=>(mark={B[4],A[...
6	9	(11=1)&(10=1)&(12=1)&(1=1)&(2=1)&(4=1)&(3=0.5)&(13=1)&(14=1)&(5=0)=>(mark={A[9]})
7	8	(1=1)&(11=1)&(2=1)&(6=1)&(12=1)&(14=1)&(3=0.5)&(7=0.5)&(8=0.5)=>(mark={A[8]})
8	6	(11=1)&(1=1)&(12=1)&(15=0)&(3=0)&(5=0)&(9=0)&(7=0)&(8=0)&(14=0)=>(mark={C[6]})
9	6	(1=1)&(11=1)&(2=1)&(6=1)&(12=1)&(10=1)&(13=1)&(4=1)&(3=0.5)&(8=0.5)&(5=0)=>(mark={A[6]})
10	6	(1=1)&(11=1)&(2=1)&(6=1)&(12=1)&(10=1)&(8=1)&(4=0)&(3=0.5)=>(mark={A[6]})
11	6	(1=1)&(11=1)&(2=1)&(4=1)&(12=1)&(10=1)&(15=1)&(5=1)&(6=0)=>(mark={A[6]})
12	5	(1=1)&(11=1)&(12=1)&(2=1)&(4=1)&(8=1)&(9=0.5)&(3=0)&(5=0)=>(mark={B[5]})
13	5	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(15=1)&(7=1)&(5=1)&(12=1)&(13=1)&(14=1)&(3=0.5)&(8=1)&(10=0)=>(mark={A[5]})
14	5	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(13=1)&(8=1)&(3=0)&(9=1)=>(mark={A[5]})
15	4	(11=1)&(1=1)&(9=0)&(12=1)&(10=1)&(2=0.5)&(3=0)&(13=0.5)=>(mark={B[4]})
16	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(14=1)&(15=1)&(7=1)&(8=1)&(5=1)&(3=1)&(9=0)=>(mark={A[4]})
17	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(14=1)&(15=1)&(7=1)&(8=1)&(3=0.5)&(5=1)&(9=1)=>(mark={A[4]})
18	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(8=1)&(7=1)&(15=1)&(3=1)&(5=1)&(9=0.5)=>(mark={A[4]})
19	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(12=1)&(10=1)&(13=1)&(8=1)&(3=0.5)&(7=0.5)&(9=0.5)=>(mark={A[4]})
20	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(13=1)&(7=1)&(9=1)&(8=1)&(15=0)=>(mark={A[4]})
21	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(10=1)&(12=1)&(15=1)&(3=1)&(7=1)&(5=1)&(14=0.5)=>(mark={A[4]})
22	4	(1=1)&(11=1)&(2=1)&(6=1)&(4=1)&(13=1)&(8=1)&(9=1)&(12=0)&(14=1)&(5=0)=>(mark={A[4]})
23	4	(1=1)&(11=1)&(12=1)&(10=1)&(6=1)&(13=1)&(15=0)&(3=0)&(5=0)&(9=0)&(4=0)&(14=0)=>(mark={A[4]})
24	4	(1=1)&(11=1)&(2=1)&(4=1)&(12=1)&(13=1)&(8=1)&(6=1)&(3=0.5)&(7=1)&(15=1)&(5=0.5)=>(mark={A[4]})
25	4	(1=1)&(11=1)&(2=1)&(6=1)&(10=1)&(12=1)&(14=1)&(8=1)&(13=1)&(9=0.5)&(3=1)=>(mark={A[4]})
26	4	(1=1)&(11=1)&(2=1)&(6=1)&(10=1)&(4=1)&(9=1)&(13=0)=>(mark={A[4]})
27	4	(1=1)&(12=1)&(6=1)&(2=1)&(13=1)&(4=1)&(3=0.5)&(5=1)&(7=1)&(15=1)&(8=0.5)=>(mark={A[4]})
28	4	(11=1)&(1=1)&(2=1)&(6=1)&(12=1)&(4=1)&(3=0.5)&(13=1)&(7=1)&(10=1)&(15=1)&(5=0.5)=>(mark={A[4]})
29	4	(11=1)&(1=1)&(2=1)&(6=1)&(3=0.5)&(9=0.5)&(4=1)&(13=1)&(5=0)=>(mark={A[4]})

IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 10, May 2010
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

36

5. Results and Discussion

Results inferred on the factors highly influencing the academic performance of students based on the analysis of data collected through the five questionnaires are illustrated here.

- Students showed genuine and incredible response in answering the questionnaire with open ended questions-with no answers to circle forcing the respondents to select an answer and make them to think through their responses.
- Hiding their identity facilitated the students to answer the questionnaire openly and honestly.
- Open-ended questions are very useful for exploring information from students even though they are difficult to process and tabulate the knowledge.
- From this study, it is inferred that academic performance is positively and significantly associated with factors such as constant care, affection, motivation, parent - student relationship, no comparison, not so strict, frequent tests, friendly, co-operation while studying, no pressure, more revision, sincere usage of study holidays before exam, special coaching, sincere advice and care by teachers, good relationship
- Academic performance is negatively associated with Factors such as less motivation, more time spent on roaming ,chatting, cell phones, wasting time with friends, spent more time on TV, more freedom, less care, disappointment, lack of affection, not so serious in studies, laziness, carelessness, lethargic, not having sound sleep, not listening in the class, less hard work, no early morning wake up, no frequent test, no good infrastructure, poor teaching, no proper notes, no special care, not having regular attendance, only last minute study are negatively associated with academic performance.

From the comparative study of the academic data collected in the year 2006 and in the year 2010 it is observed that factors influencing academic achievement doesn't change much over time.

It will be appropriate to quote four things found in verses 24 – 28 of Proverbs 30 from the Holy Bible related to wisdom given in table 6.

Table 6. "Four things on earth are small yet they are extremely wise"

1	**Ants** are creatures of little strength yet they store up their food in summer;	Future Planning and hard work	Ant plans properly and it can lift and pull things 50 and 30 times of its own weight respectively to face the future[5].
2.	**Coneys** are creatures of little power, yet they make their home in the crags;	Protection	Coneys have little or no tail and have rump that easily falls away when seized by a predator.
3	**Locusts** have no king, yet they advance together in ranks.	Co-ordination, partnership / good friends	Can cover one fifth of earth's land surface. A desert locust swarm can be 1,200 square kilometers in size and pack between 40 and 80 million locusts into less than half a square mile(one square kilometer [5].
4.	A **spider** can be caught by the hand, yet it is found in king's palaces.	Aim high, set goal, Positive and motivated thoughts	Spider silk is stronger than bone and twice as elastic as nylon. Many experts claim that the silk of an orb weaver spider is so strong that a strand as thick as a pencil could stop a jumbo jet in flight [5].

Blessings and curses are set before us in this world. The choice is ours. The flesh lusts against the spirit and the spirit against the flesh and these are contrary one to the other. From the results extracted in our study, it could be realized that those who

- Aim high like **Spider**
- Plan their activities like **ant**
- Anchor themselves to protect from the storms of physical, psychological, emotional, temptations especially students protected from the temptations of TV, movie, enjoyment, chatting, roaming, laziness, lack of confidence, lack of motivation etc like **Coneys**

- Work smart like **Locust** with the co-operation and co-ordination of friends and with the motivation of parents, teachers could help them attain their scholastic goals fruitfully.

6. Conclusions

Identifying the influence of various factors affecting students' academic performance is of great importance to educators and parents. Factors indentified and guidance rules derived in this study from high and low achieves data will be more helpful to students who have difficulties in their scholastic performance. This study will facilitate the students to achieve their

performance systematically and to have an awakening about their area of deficiency to be strengthened to have the best performance. Data collection is an important and a challenging task in data mining. Good quality data plays a vital role in this task. Therefore it is important to ensure that genuine and good quality data is used in the analysis. This study also addresses the problem in the process of collection of reliable and stable data which will be helpful for any research on data analysis. We plan to have an extended study to check whether these rules changes with time, place and with volume of data.

Bibliography:

[1] Behrouz Minaei – Bidgoli, Deborah A.Kashy, GerdKortemeyer and William F.Punch. "Predicting student performance:An Application of data mining methods with the educational web based system LON – CAPA" , 33rd ASEE / IEEE Frontiers in Education Conference 2003, 0-7803-7444-4/03/2003 IEEE

[2] Daniel Delic, Hans-J.and Lenz, and Mattis neiling, "Improving the Quality of Association rule mining by means of Rough sets",First International Workshop on soft Methods in Probability and Statistics SMPS 2002.

[3] Daniza M.Ivanovic , Boris P. Leiva , Hernana Perez , Manuel G. Olivares , Nora S. Diaz, Maria Soledad C. Urrutia , Atilio F. Almagia, Triana D. Toro , Patricio T. Miller , Enrique O. Bosch and Cristian G. Larraín "Head size and intelligence, Learning, nutritional status and Brain development Head, IQ, learning, nutrition and brain". Neuropsychologia, Vol.42, No.8, 2004,pp.1118 – 1131.

[4] Gintautas Dzemyda "Multidimensional data visualization in the statistical analysis of curricula", Computation statistical and Data Analysis Vol.49, No.1, 2005,pp.265 – 281.

[5] http://www7.natioanlgeographic.com/ngm/data/2001/08/01/nationalgeographic.com,http://en.wikipedia.org/wiki/user:reywas92/trivia

[6] Blaženka Divjak and Dijana Oreški."Prediction of Academic Performance Using Discriminant Analysis", 31st Int. Conf. on Information Technology Interfaces, June 22-25, 2009.

[7] Jacek Sienkiewicz "Rough sets for Boolean functions Minimization Research Report of Institute of Computer Science", Warsaw University of Technology,Electronic Bulletin of the Rough Set Community EBRSC Feb 2,1995, Vol.3, No.2.

[8] Keyun Hu , Yuchang Lu and Chunyi Shi "Feature Ranking in Rough Sets" International Journal of Computer and Information Sciences Vol.11, No.5, 1982,pp.341 – 356.

[9] Kristina J.Andren "Characteristics of Higher Performing,HigherPovertyElementary chools".Paper Presented at the Annual Meeting of New England, Educational Research Organization , Portsmouth,New Hampshire April 2003.

[10] Mevlüt TÜRE, Zekeriya AKTÜRK,Imran KURT,Nezih DAĞDEVIREN "The Effect of Health status , Nutrition and some other factors on low School Performance using Induction Technique".Trakya Univ Tip Fak Derg Vol.23, No.1, 2006,pp.28 – 38.

[11] Rahel Bekele and Wolfgang Menzel "A Bayesian Approach to predict performance of a student" (BAPPS): A case with Ethiopian students, IASTED International Conference on Artificial Intelligence and Applications, part of the 23rd Multi-Conference on Applied Informatics,2005,pp.189-194.

[12] Salazar A, J. Gosálbez, I. Bosch, R.Miralles, L.Vergara, "A case study of knowledge discovery on Academic Achievement, Student desertion and student retention", In.preceedings of International conference on Information Technology: Research and education, 2004. ITRE 2004.July 2004 pp150-154.

P.G.JansiRani, She has completed her M.Sc (Applied Mathematics) and M.Phil (Mathematics) through Bharathidasan University. Now she is a faculty in the department of Mathematics at Sethu Institute of Technology, Madurai, Tamilnadu, has16 years of teaching experience in Engineering college. Research scholar at School of Mathematics, under the guidance of Dr.R.Bhaskaran, Madurai Kamaraj University, Madurai, Tamilnadu, India. She has published 1 paper in international journal and presented 1 paper in international conference.
Fields of interest: Data Mining and Rough set theory, Operations Research.

R.Bhaskaran did his M.Sc., at IIT Chennai and Ph.D at the Ramanujan Institute for Advanced study in Mathematics. He joined the Madurai Kamaraj University in 1980, as a lecturer. At present, he is Sr,Professor. His areas of interest are Non-archimedean functional analysis, Image processing, Data Mining, Software development for learning Mathematics.