A Data Mining Based Approach to Evaluate Assessment Performances of Graduating Students of Schools

Rituraj Jain[1], Yohannes Bekuma[2], Bipin Pandey[3], Jaishri Gothania [4]

[1,2] Department of Electrical and Computer Engineering, Wollega University, Nekemte, Ethiopia.

[3] Department of Computer Science & Engineering, Dronacharya Group of Institutions, UP, India.

[4] Department of Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India.

ABSTRACT

In the current trends of advance computing methodologies, data of students' performances in different grades can be used to improve the quality of managerial decisions. Student's academic performance is based upon various factors like personal characteristics and psychological factors. Educational database contains useful information for predicting a students' performance, rank factor and details. By applying different data mining techniques to educational data to analyse them as well as to develop good methods to knowledge gain and management. Finding better correlation between different data variables can allow us to make better and beneficial decision which can facilitate better resource utilization in terms of educational service delivery.

This paper aims to analyses and predict the correlation between English, Mathematics and science subjects in terms of student academic result in 10th and 12th grade by using Aprior data mining techniques which mines required information. National level examination results of 10th and 12th grade students' have been used for this research. The results show strong relationships between subjects as well as subject relationships with gender of the student in a specific grade. The results of this research help educationist to develop proper education model to improve results and to get better achievements in the areas where lacking.

Introduction

The national examination is one of the forms of evaluation done on a national scale in the education world and adjusted with the national achievement standards. The national exam is carried out at high school at grade 10 and preparatory school levels on 12th grade to measure student graduate competence nationally in certain subjects as well as to map the level of achievement for students at the school level and the regional level.

Very useful hidden patterns from the large dataset can be identify by using data mining. By applying data mining techniques on educational database, we can find unknown patterns for designing a better system which in turn lead to improve student's performance in learning as well as in assessments. Reasons of good or bad performances belong to the main interests of teachers, because they can plan and customize their teaching program, based on the feedback of policy makers. Hence, in this paper we examined the extent to which the academic achievement of 12th grade students is related to achievement of 10th grade and association between sciences, English and mathematics subjects by using Aprior

algorithm. This study is tried to build a model and addressed the factors associated with students' academic achievement by applying data mining Techniques on the national educational assessment and examination data set taken from schools.

Data mining or knowledge discovery refers to the process of finding interesting information in large repositories of data. Data mining process can also be defined as process of exploring patterned based new knowledge by applying algorithms. These explored patterns are then analysed yielding knowledge [1]. Organization's decisions can be made based on trends and behaviours predicted by data mining tools [2]. Based on the past history-based data collection, data mining based automated analysis provides better prediction of future trends and behaviours. Data mining tools find hidden patterns and relevant information which in-turn may help to answer typical business-oriented questions which are more time consuming to resolved in traditional manner [3].

LITERATURE REVIEW

Many researchers have been discussed the different factors that affects the student academic performance in their research. External and internal classroom-based factors or characteristics affect the students' academic performance. Student's competence in English, Class factors like schedule, size, facilities, environment, homework, test results are some of the internal factors. Other than these internal factors some of the external classroom factors also affects student's performances. These external factors are like extracurricular activities, family background, social environment and other problems. According to his research studies students' performance depends on many factors some of them are facilities adopted for learning, age factor and gender etc. [4].

Many of the researchers used the GPA, result of particular subject or the previous year result, students' academic activities managing strategies and background characteristics (i.e., family income, parents' level of education, guidance from parents and different negative scenarios at the home) were indirectly related to their composite scores in high school [5],[6]. Proper guidance to students is also one of the important factors which help students to improve study attitude and habits which is also lead to get good academic achievement. The students who are properly guided by their parents have performed well in the exams. Student performance also have effects of guidance from the teacher and parents.

Lot of arguments have been seen to highlight gender-based achievements in math science subjects. But many of the research also shown that this difference has been shorten with time in last few decades. This shows that women exposure and increased interest in maths and science subjects [7]. Still many researches also putting this outcome in dilemma by showing difference in academic achievements based on gender difference in maths and science subjects.

As per the Minister of Education's report on the New Zealand school system showed that in School Certificate examinations in male performance was better than in maths subject while female performance was better than male in science subjects. The researcher used methods including parental interviews, teacher assessments, standardized testing and interviews with the children as data collection and he found that if boys and girls behaved in similar ways in the classroom setting there would have been no evidence of consistent gender related differences in school achievement [8].

Gender, age and finance are not significant predictors of academic performance and there was no significant difference in academic performance based on age, gender and financial status rather the student's character and behaviours are basic and foremost determinants of academic performance. So, the researchers recommend that counselling centers should open to handle varying problems [9].

According to the study to investigate factors that predict student performance of Debre Markos University students Models were built and tested by using a sample dataset of 11,873 regular undergraduate students. He used decision tree and Bayes as a classification technique and the generated rules were studied and evaluated. Data collected from MS_EXCEL files, and it has been pre-processed for model building. He obtained from his research result that Higher Education Entrance Certificate Examination result of a student, and sex were the main determining attributes to classify university students' academic performance as failure/success [10]. Likewise, another study conducted to Analyses determinants Influencing Students' Academic Achievement in English and Mathematics subjects focused on ending classes of schools (grade 8, 10 and 12) at Bahir Dar and Debre Tabor Towns at both government and non-government schools in Ethiopia. The result showed that, on average, the academic achievements of female students were better than male students in mathematics at grade eight, ten and twelve [11].

Educational data mining is a study that make use of statistical, machine-learning, and data-mining algorithms over the different types of educational data in order to analyse these types of data to resolve educational research issues. Even if there are many approaches, the researcher applied Bayesian Network classification methods up on student database to predict the students' academic performance. The result of the study will help the students to improve their performance and also it helps teacher to identify those students which needs a special attention to reduce failing ratio and taking appropriate action at right time. Based on the Experimental Results he concludes that AODEsr Algorithm predict more accuracy than any other Algorithms [12].

Educational data mining (EDM) is a field that exploits machine-learning, statistical and data-mining algorithms over the different types of educational data. Analysis of data has to be done with the objective of explore the unique educational data settings to resolve educational research issue. It will also help to better learning environment and computational approaches by understanding the learners [13].

Classification, one of the most common data mining methods, seems to be a human imperative. Any classification technique can be used for training-based prediction where variables to be predicted are known with historical data. The historical data is used to build a model that explains the current observed behaviour. When this model is applied to current inputs, the result is a prediction of future behaviour [14].

Data mining is the technique used for extracting useful knowledge form the historical large amount pf data stored in data warehouses and repositories. Classification clustering, and association rule mining are some of task or techniques of data mining used for extracting meaningful and interesting knowledge. Relationship between data items will shows by using association rules.

Association rule generation consists of two separate steps which are applying minimum support followed by applying minimum confidence constraints. Frequent items set find by applying minimum support and then rules will be formed by using minimum confidence. Support & confidence are the normal method used to measure the quality of association rule. Support for the association rule X$\rightarrow$Y is the percentage of transaction in the database that contains XUY. Confidence for the association rule is X$\rightarrow$Y is the ratio of the number of transactions that contains XUY to the number of transactions that contain X [15]. Association rule can be used in educational data mining and teacher's evaluation system for analysing the learning data. Student's performance is evaluated using association rule mining algorithm. The research has been done on assessing student's performance based on various attributes graduation mark, test mark, attendance and university result. In this study important rules were generated to measure the correlation among selected attributes which would help to improve the student's academic performance [16].

The classification method applied as data mining technique to evaluate student "performance, they used decision tree method for classification. They applied the association rule mining analysis based on students' failed courses to identify students' failure patterns. A relationship has been trying to identified between failed courses and causes for failure to enhance the performance of the low graded students [17]. Bayesian Classification Method also applied as a data mining technique and concluded that senior secondary results, place of living, medium of study, mother's qualification, students' other habits, annual income of house and family status were the factors affecting student academic performance [18]. Another study conducted on the student performance using association rule technique and they find the interestingness of student in opting class teaching language [19].

METHODOLOGY

Primary data used for this research is the national educational assessment and examination data set taken from schools. Secondary data for instance reviews important document to gain farther information related with student achievement. We have used the CRoss-Industry Standard Process for Data Mining (CRISP-DM) model, which contains six phases. For applying data mining on the general problem-solving strategy of business unit, CRISP-DM (CRoss-Industry Standard Process for Data Mining), a non-proprietary and freely available standard process used. The CRISP–DM demands that data mining be seen as an entire process, from communication of the business problem through data collection and management, data pre-processing, model building, model evaluation, and finally, model deployment as shown in figure 1.

In the first phase understanding the project objectives and requirements from a business perspective will be identified. Then after understood knowledge will be convert into a data mining problem definition with preliminary plan to achieve the objectives. The source of data for this research taken is the national educational assessment and examination data set of 10th and 12th grade taken from schools.

Next phase starts with collection of data and followed by activities to understating the data, data quality problems and detecting interesting

subsets. Education data set contains details information on student name, address of student, student academic result of 10th and 12th grade, school type, age, gender, teachers' qualification, availability of media technology at school etc.

Data preparation is the most important phases of the data analysis activity which involves the construction of the final data set (data that will be fed into the analysis tool) from the initial raw data. Original data sets will be converted in smaller data sets which can significantly improve the efficiency of data mining. This task includes: attribute selection, filling missed values, correcting errors, or removing outliers (unusual or exceptional values), resolve data conflicts using domain knowledge or expert decision to settle inconsistency. Since there is no missing value occurred on the data set, no missing value replacement technique applied on this research. Collected excel format prepared in some Weka understandable formats. Then pre-processing activities are performed and the file is saved into Weka acceptable comma separated values (CSV) or comma delimited file format. The CSV file format is converted into Attribute Relation File Format (ARFF) by using Weka mining software, to take advantage of easier data manipulation and also compatible interaction with Weka software.
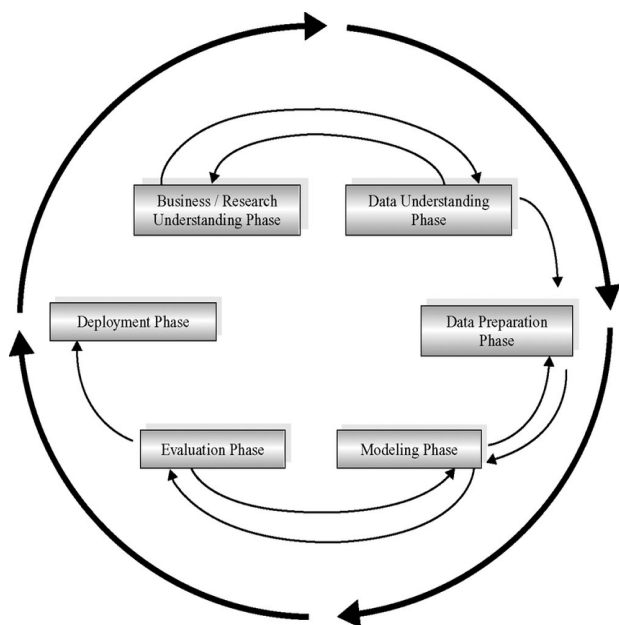

Figure 1: Phases of the CRISP-DM reference model Source [2]

During scan of the pre-processed data some basic statistics summary was produced for each attribute

into a form acceptable by the selected data mining software Weka.

In this phase various modelling techniques are selected and applied, and their parameters are adjusted to optimal values. As it is indicated previously, the purpose of this research is to develop association rules model. Association rule data mining technique is applied in predicting correlation between 10th and 12th grade students' academic result. WEKA allows the resulting rules to be sorted according to different metrics such as confidence, leverage, and lift. In this research, we have selected lift as the criteria entered certain value as the minimum value for lift (or improvement) is computed as the confidence of the rule divided by the support of the right-hand-side (RHS). In a simplified form, given a rule L => R, lift is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e. lift = $Pr(L,R) / Pr(L).Pr(R)$.

If this value is 1, then L and R both are independent. The higher this value, the more likely the existence of L and R together in a transaction is not just a random occurrence, but because of some relationship between them. Defaults value of rules will be 20 so that program will report only top 20 rules. The upper bound and lower bound for minimum support is set to 1.0 (100%) and 0.1 (10%) respectively. The algorithm starts with upper bound support and then decrease support incrementally. It will be continued until either the specified numbers of rules are generated, or the lower bound for min. support is reached. The significance testing option is only applicable in the case of confidence and is by default not used (-1.0)**.** With the given values of frequency count of each LHS, RHS as well as values of lift, confidence, leverage and threshold values of support and lift, rules will be discovered. By this stage, Models obtained are evaluated in details and reviews also carried to check achievements of business objectives. Evaluation is the key to making real progress in data mining. Once the model is designed, results and their significance will be evaluated that is we have to evaluate mined patterns to check their effectiveness on the system. The evaluation process is also carried out to identify interesting patterns representing knowledge based on some lift as an interestingness measure.

ASSOCIATION RULE IN WEKA

There are a number of machine intelligent techniques that are available in the market but at the same time not all tools are the best for all problems in the data set. Based on the algorithm used different results will be produced for different data sets. In this study we are testing one algorithm by adjusting different parameters value based on Apriori method association rule. Our aim is to find the best model. The association rule mining can be considered as a two-step process which are: find all frequent itemsets and generate strong association rules from the frequent itemsets [16].

Association rule learning is process known as finding interesting relations between variables in large databases and discover strong rules in databases using interesting measures. Which means in association we are interested in finding out association between items with no particular focus on a target one, whereas in classification we basically map the set of record (attributes, variables) into the class attribute.

Correlations between items in a dataset can be find using association rule mining. It also used to discover interesting relationships from a student's academic result, as well as relationships between each pattern of subjects. If-then statements concerning attribute values produced based on the relationships discovered by association rules mining [20].

DATA PREPARATION AND PRE-PROCESSING

It is well known that success of every data mining algorithm is strongly dependent on a quality of the data used for mining. Data pre-processing task could be critical and a very complicated task. The purpose of data pre-processing is to clean the noisy data, extract and merge the data from different sources, and then transform and convert the data into a proper format [2].

5.1 Description of The Original Data

Pre-processing is done on the national educational assessment and examination data set of 10th and 12th grade taken from schools.

Table 1 is showing different attributes of raw data and their data types. Raw data extraction and target data selection for study is done as the first steps in the processing. Subset of the database to selected from the large dataset for knowledge discovery. Data mining tools can be used on selected data set to extract meaningful information and patterns from the raw data.

Attribute		Description	Type	Width
Student_id		Student's Registration Number	double	10
Name		Student's Name	text	25
Sex		Student's Gender	char	1
Age		Student's Age	int	2
Nationality		Student's Nationality	text	20
Disability		Student's Disability	text	10
School_code		Student's School Code	text	12
School_name		Student's School Name	text	25
ICT		ICT Technology Availability in School	boolean	1
Staff_Qualification		Minimum Academic Qualification of Teaching Staff in School	text	2
Subject	Amharic	Results of Amharic	int	3
	English	Results of English	int	3
	Physics	Results of physics	int	3
	Chemistry	Results of chemistry	int	3
	Biology	Results of biology	int	3
	Mathematics_NS	Results of natural science mathematics	int	3

	Mathematics_SS	Results of social sciences mathematics	int	3
	Civic_Ethics	Results of civic and ethical education	int	3
	Geography	Results of Geography	int	3
	History	Results of History	int	3
Total		Total Sum of Marks of All Subjects	int	3
Unique_ID_Code		Unique Number given to Each Student Card	float	10

Table 1: Set of Partial Attributes of the Original Dataset

5.2 Attribute Selection

To increasing the efficiency of any data mining algorithm task, it is important to select relevant attribute features. The dataset used from national examination and assessment agency has many attributes, from which we have to extracts all relevant attributes. Following attribute are selected for data mining processing:

data set contains many attributes and to decide on the relevant attributes for this study, we have discussed with domain expert in the area. As described in table, the following attributes are selected: Student-id, Sex, Age, School code, school's media technology, teachers' academic qualification, English Physics Chemistry Bio Maths 10th and 12th grade marks and Total. Main objective of this research is to make a proper association between different attributes. So, only the attributes related to student's performance have been used. By excluding the irrelevant attributes data mining task will become simple and give more accurate results. Total 17 attributes have been extracted as relevant attributes for analysis. Table 2 shows the details of the relevant attributes selected for building and test the model.

Attribute		Description	Type	Width
Student_id		Student's Registration Number	double	10
Sex		Student's Gender	char	1
Age		Student's Age	int	2
Disability		Student's Disability	text	10
School_code		Student's School Code	text	12
ICT		ICT Technology Availability in School	boolean	1
Subjects	English_10	Results of English (10th Grade)	int	
	Physics_10	Results of physics (10th Grade)	int	3
	Chemistry_10	Results of chemistry (10th Grade)	int	3
	Biology_10	Results of biology (10th Grade)	int	3
	Mathematics_10	Results of mathematics (10th Grade)	int	3
	English_12	Results of English (10th Grade)	int	3
	Physics_12	Results of physics (10th Grade)	int	3
	Chemistry_12	Results of chemistry (10th Grade)	int	3
	Biology_12	Results of biology (10th Grade)	int	3
	Mathematics_12	Results of mathematics (10th Grade)	int	3
Total		Total Sum of Marks of All Subjects	int	3

Table 2: Relevant Attributes Extracted

5.3 Creating New Table

New table is created to summarized and aggregate 10th grade and 12th grade information of each students without losing the relevant data. It is required to compare dataset of 10th grade and 12th grade so that proper merging of data can be done on the corresponding field values by comparing the students name and other attributes.

5.4 Discretizing Data

Discretizing of different data items like numerical or continuous attributes is required as association rule mining is only applicable to categorical data. It is also performed to increase the interpretation and comprehensibility. Data is divided into categorical classes in discretization for make them more understandable. Other than student reg.no all other numerical data are discretized and separated in new table. Simple binning methods used for 12th and 10th grade subjects' marks and the equal-width method for the others attribute. The labels that we have used in the mark attribute are A, B, C for good, average, under average respectively.

5.5 Data decoding and Attribute Transformation

Data decoding is the transformation of data in the forms which are required to use for prediction by the model. Some of the attributes used for mining process are of numerical types so that these attributes are required to be converted into its representation or columns. For example, "school code" is the attribute which also defines the following information "student's region, exam center, school's name and school's type". So, it is represented 8 alphanumerical codes. These 8 alphanumerical codes are converted into their respective symbolic values. The following table 3 shows the sample of such data decoding.

| School_code | Reformated_Values | | | |
	Regional_Data	Zonal_Code / City_Name	Type_of_School	School_No
110010124	08 (oromia	Zonal 08 /	07 (Private)	24

| | region) | Nekemte | | |

Table 3: Actual Values of each registration code and its value represent

IMPLEMENTATION AND RESULT ANALYSIS

For data mining of education data set and to get more insight in learning data, association rules are proved to be very useful. It takes thresholds for support and confidence parameters as well as also take measurements to keep interesting and meaningful rule by filtering the unwanted rules. In this research good association rules are generated by altering the parameters of Apriori algorithm. The reliability of the rule will be measure by the lift. If rule value is less than 1 then it will be rated as uninteresting by the lift. Based on the formulation and derived limitations, the differs arise between confidence and lift. Confidence is sensitive to the probability of consequent (Y). Higher frequency of Y will ensure a higher confidence value even if there is no true relationship between X and Y. This problem can be avoided by increasing the threshold but then also we may have problem losing interesting pattern having less frequency. In contrast to confidence, lift is not vulnerable to the rare items problem. It is focused on the ratio between the joint probability of two item sets with respect to their expected probabilities if they are independent. The upper bound for minimum support is set to 1.0 (100%) and the lower bound to 0.1 (10%). The algorithm starts with upper bound support and then decrease support incrementally. It will be continued until either the specified numbers of rules are generated, or the lower bound for min.

6.1 Correlation Between Subjects (English, Math, Physics, Chemistry, Biology) in 10th Grade

Association Rule Extracted Between Subjects	Confidence Value	Lift Value	Leverage Value	Conviction Value
Mathematics_10 =B =>	0.31	1.03	0	1.01

Biology_10 = A					Chemistry_10 = A => Biology_10 = A	0.68	2.26	0.1	2.18
Mathematics_10 =B => Biology_10 = B	0.36	1.08	0.01	1.04	Physics_10 = C=> Chemistry_10 =C	0.31	1.07	0.01	1.03
Mathematics_10 =C => Chemistry_10 = B	0.38	1.08	0.01	1.05	Physics_10 = C => Biology_10 = C	0.33	1.05	0.01	1.02
Mathematics_10 = C => Chemistry_10 = C	0.39	1.34	0.04	1.16	Physics_10 = C => Biology_10 = B	0.36	1.08	1.01	1.04
Mathematics_10 =C => Biology_10 = C	0.39	1.26	0.03	1.13	Physics_10 = B => Chemistry_10 = A	0.38	1.37	0.04	1.17
Mathematics_10 = B => Physics_10 = B	0.39	1.02	0	1.01	Physics_10 = C=> Chemistry_10 =B	0.38	1.08	0.01	1.04
Mathematics_10 = B => Chemistry_10 = B	0.4	1.15	0.02	1.09	Physics_10 = B => Biology_10 = A	0.39	1.31	0.04	1.15
Mathematics_10 = C => Physics_10 = C	0.46	1.09	0.01	1.07	English_10 = C=> Chemistry_10 = C	0.37	1.27	0.01	1.13
Mathematics_10 =A => Biology_10 = A	0.66	2.2	0.06	2.07	English_10 = B=> Chemistry_10 = B	0.37	1.04	0.01	1.02
Mathematics_10 = A => Chemistry_10 = A	0.69	2.5	0.07	2.34	English_10 = C => Biology_10 = C	0.42	1.35	0.04	1.19
Chemistry_10 = B => Biology_10 = B	0.46	1.31	0.04	1.21	English_10 = C=> Physics_10 = C	0.45	1.5	0.01	1.04
Chemistry_10 = C => Biology_10 =C	0.48	1.53	0.05	1.32	English_10 = C => Mathematics_10 = C	0.46	1.2	0.03	1.14
					English_10 = B=> Physics_10	0.48	1	0	1

= B				
English_10 = B => Mathematics_10 = B	0.5	1.06	0.01	1.04
English_10 = B => Biology_10 = B	0.51	1.07	0.01	1.07
Biology_10 = A => English_10 =B	0.55	1.35	0.04	1.16
English_10 = B => Biology_10 = A	0.55	1.15	0.02	1.16

Table 4: Association rule on 10th grade science subjects

By analysing the table 4 and respective bar chart shown in figure 2, we can find some of the following facts:

1. The probability of students who score "A" in chemistry have got "A" in Biology is higher.
2. The probability of student who get "B" in chemistry cannot get "C" in biology.
3. The probability of student who scores "A" in chemistry does not score "B" or "C" in biology.
4. Students who score "A" in english do not have any association with other subject's result but the english result score "B" have some association with other subject's result.

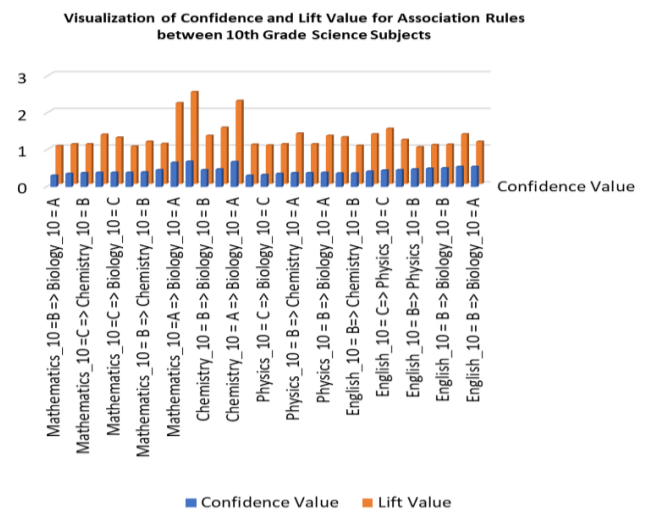5. There is strong relationship between results of mathematics with both biology and chemistry result.

Figure 2. Bar Chart representation of Confidence and Lift Values for association rules between 10th Grade Science Subjects.

6.2 Correlation Between Subjects (English, Math, Physics, Chemistry, Biology) in 12th Grade

By analysing the table 5 and respective bar chart shown in figure 3, we can find some of the following facts:

1. Both biology and chemistry subjects' results have better correlation than other subjects.
2. Both maths and english subject's result have a better correlation with biology subject result.
3. Maths and english subjects have more effects on study of biology subject.

Association rule extracted between subjects	Confidence value	Lift value	Leverage value	Conviction value
Mathematics_12= C => Biology_12 = C	0.17	1.23	0.02	1.04
Mathematics_12 = C => Chemistry_12 = C	0.3	1.25	0.04	1.08
Mathematics_12= B=> Biology_12 = A	0.62	1.67	0.06	1.66
Biology_12= A => Mathematics_12 = B	0.74	1.67	0.06	1.27
Physics_12 = C => Mathematics_12 = C	0.8	1.08	0.05	1.31
Biology_12= C => Mathematics_12 = C	0.91	1.67	0.06	2.95
Mathematics_12 = C => Physics_12 = C	0.91	1.08	0.05	1.17
Chemistry_12 = "C => Mathematics_12 = C	0.93	1.25	0.04	3.57

English_12 = C => Mathematics_12 = C	0.78	1.05	0.03	1.1
Mathematics_12 = C => English_12 = C	0.71	1.05	0.03	1.12
English_12 = C => Physics_12 = C	0.86	1.03	0.02	1.17
Physics_12 = C => English_12 = C	0.69	1.03	0.01	1.06
English_12 = A => Chemistry_12 = A	0.79	1.90	0.01	1.03
Chemistry_12 = C => English _score = C	0.73	1.09	0.01	1.02
English_12 = C => Chemistry_12 = A	0.61	1.01	0	1.02
Chemistry_12 = B => English_12 = C	0.68	1.09	0	1.02
English_12 = B => Biology_12 = A	0.73	1.26	0.03	1.18
Biology_12 = A => English_12 = B	0.39	1.26	0.03	1.13
English_12 = C => Biology_12 = C	0.76	1.13	0.01	1.37
Chemistry_12 = A => Biology_12 = A	0.75	1.96	0.06	2.31
Biology_12 = A => Chemistry_12 = A	0.61	1.96	0.06	1.22
Physics_12 = C => Chemistry_12 = C	0.27	1.13	0.03	1.04
Chemistry_12 = C => Physics_12 = C	0.94	1.08	0.01	2.9
Physics_12 = C => Biology_12 = C	0.15	1.11	0.01	1.02
Biology_12 = C => Physics_12 = C	0.15	1.11	0.01	1.02

Table 5: Association rule on 12th grade science subjects
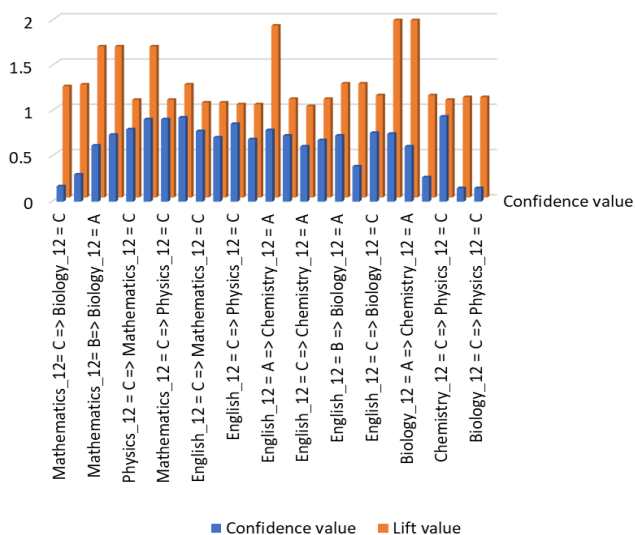

Figure 3. Bar Chart representation of Confidence and Lift Values for association rules between 12th Grade Science Subjects.

6.3 Association of 10th And 12th Grades Student's Result and Modelling

Table 6 shows association rules extracted by Apriori algorithm on 10th and 12th grade student results. To select most appropriate and relevant attributes to extract effective association rule, CfsSubsetEval filtering is used. According to this 10th grade chemistry results of student as well as 12th grade english and biology subject's results of student are selected with the 12th grade total marks. Discovered information from the design model can help teachers to take proper decision based on examination performance to improve learning process of any student. Where some of the rules which are not much interesting and relevant, some of the rule patterns shows unexpected relationships which can be used by educationist to improve student's learning activities.

Subject	Association rule extracted	Confidence value	Lift value	Leverage value	Conviction value
Physics	Physics_10 = B => Physics_12 = C	0.76	1.04	0.01	1.11
	Physics_10 = C => Physics_12 = C	0.78	1.05	0.01	1.08
Biology	Biology_10 = A => Biology_12 = A	0.85	1.24	0	1.06
	Biology_10 = B => Biology_12 = B	0.85	1.27	0	1.06
	Biology_10 = C => Biology_12 = B	0.89	1.27	0	1.04
Chemistry	Chemistry_10 = C => Chemistry_12 = C	0.84	1.14	0.01	1.07
English	English_10 = C => English_12= C	0.47	1	0.05	1.78
	English_10 = B => English_12= B	0.52	1.01	0.03	1.19
Mathematics	Mathematics_10 = C => Mathematics_12 = C	0.53	1.02	0.02	1.18
	Mathematics_10 = B => Mathematics_12 = C	0.54	1.02	0.01	1.18

Table 6: Association rules extracted by Apriori algorithm on 10th and 12th grade student results.

Any student has scored grades "B" and "A" in english and chemistry respectively in 10th grade then there is more probability to score "proficient" results in total marks in 12th grade. From the finding it can also see that a student may get "basic" results in 12th grade if he/she scored minimum "C" in english and maths and scored "B" in biology subject.

6.4 Evaluation

In this research we trying to evaluate result of 12th grade of student by analysing and making correlation with exam result of 10th grade. For this, we have used the confidence and lift value between antecedent and consequent. For example, the first rule with lift is 1.3 means there is a positive correlation between the antecedent 10th grade student biology results and the consequent total mark on 12th grade student. Importance of the rule can be measure by the lift value.

CONCLUSION

The result of the male students in English, Mathematics, Physics, Chemistry and Biology on both 10th and 12th grade have a better correlation than female students. Student's result of 12th grade was dividing in to three standards as: C (less than 55), B (between 55 and 65), proficient (more than 70). By analysis it can be concluded that around 1.5% were scored above basic, 50.5 % achieved basic and 48% were below basic. This indicates that most student still lower academic performance.

It is also analysed that there exist positive relationships between the english, maths, chemistry, physics and biology in both 10th and 12th grade. This shows that if performance of a student is good in subject then he/she also perform the same in the other subject. Looking at instructional language English and other subjects it was the highest with biology (English= A => Chemistry= A, con=0.79 and lift=1.90). Subjects such as biology when compared with other subject are more influenced by language ability. The same like chemistry and biology are strongly correlated each other.

REFERENCES

[1] Maimon, Oded & Rokach, Lior (2010). Introduction to Knowledge Discovery and Data Mining in Data Mining and Knowledge Discovery Handbook, Springer US, ch. 1, pp.1-15, 2010 DOI: 10.1007/978-981-15-5232-8_12

[2] Daniel T. Larose, Chantal D. Larose (2014), Discovering Knowledge in Data - An Introduction to Data Mining, 2nd ed Wiley Series on Methods and Applications in Data Mining, IEEE Computer Society Wiley, ch 1, pp 1-14, 2014

[3] Deshpande, Shrinivas & Thakare, V. M. & Mandal, H & India, Amravati. (2010).

Data Mining System and Applications: A Review. International Journal of Distributed and Parallel systems. 1. 10.5121/ijdps.2010.1103.

[4] Hussain, A (2006). Effect of Guidance Services on Study Attitudes, Study Habits and Academic Achievement of Secondary School Students, Bulletin of Education and Research Vol. 28, No.1 pp. 35-45 2006

[5] Naqvi, Sayyed (2006), Factors Affecting Students' Performance A Case of Private Colleges, Bangladesh e-Journal of Sociology. Vol 3 No. 1, 2006

[6] Noble, Julie P, Roberts, William L. and Sawyer Richard L. (2006), Student Achievement, Behavior, Perceptions, and Other Factors Affecting ACT Scores, ACT Research Report Series 2006 – 1, 2006.

[7] Halpern, D.F., LaMay, M.L (2000). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. Educational Psychology Review 12, 229–246 (2000). https://doi.org/10.1023/A:1009027516424

[8] Fergusson DM, Horwood LJ(1997). Gender differences in educational achievement in a New Zealand birth cohort. New Zealand Journal of Educational Studies, 1997; 32(1): 83-96

[9] E.E. Ebenuwa-Okoh (2010) Influence of Age, Financial Status, and Gender on Academic Performance among Undergraduates, Journal of Psychology, 1:2, 99-103, DOI: 10.1080/09764224.2010.11885451

[10] Yehuala, Muluken Alemu (2015). "Application of Data Mining Techniques for Student Success And Failure Prediction (The Case Of Debre_Markos University)." International Journal of Scientific & Technology Research 4 (2015): 91-94.

[11] Endris Assen Ebrahim, Mulugeta Alemayehu Dagnaw (2015), Analysis of Determinants Influencing Students' Academic Achievement in English and Mathematics Subjects of Ending Classes of Schools at Bahir Dar and Debre Tabor Towns, Ethiopia, International Journal of

Secondary Education. Vol. 3, No. 6, pp. 61-66, 2015 10.11648/j.ijsedu.20150306.12

[12] Sundar, Praveen. (2013). A Comparative Study for Predicting Student's Academic Performance Using Bayesian Network Classifiers. IOSR Journal of Engineering 2250-3021. 3. 2250-3021. 10.9790/3021-03213742.

[13] Vaibhav P. Vasani1, Rajendra D. Gawali, Classification and Performance Evaluation Using Data Mining Algorithms, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, no 3, 2014

[14] Berry, M. and G. Linoff. Data mining techniques - for marketing, sales, and customer support 2nd ed., Wiley computer publishing, ch 1, pp 8-9, 1997.

[15] Han, Jiawei & Kamber, M. & Pei, J., Data Mining: Concepts and Techniques, 2nd ed, The Morgan Kaufmann series in data management systems, ch 6, pp. 246 -250, 2006

[16] Borkar, Suchita & Kannan, Rajeswari, Predicting Students Academic Performance Using Education Data Mining, International Journal of Computer Science and Mobile Computing, Vol. 2, no 7, pp. 273-279, 2013

[17] Anuradha, C & T, Velmurugan. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. Indian Journal of Science and technology. Vol 8 (15). 10.17485/ijst/2015/v8i15/74555.

[18] M., Durairaj, C. and Vijitha, Educational Data mining for Prediction of Student Performance Using Clustering Algorithms, International Journal of Computer Science and Information Technologies, Vol. 5, no 4, pp 5987-5991, (2014).

[19] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, Data mining model for higher education system, Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010

[20] Kumar,V, Anupama,C., Mining Association Rules In Student's Assessment Data, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5 No 3, pp 211-216, 2012.