COURSE GRADE PREDICTOR USING ASSOCIATION RULE MINING

ABDUL WAHAB[1], WAJEEHA FAROOQI[1] , ADNAN ABID[2],
AMJAD IQBAL[3] ,FAROOQ AHMAD[4] , AND MUHAMMAD SHOAIB
FAROOQ[5]

[1]Department of Computer Science & information Technology
abdul.wahab@cs.uol.edu.pk
[1]Department of Computer Science &information Technology
wajeeha.farooqi@cs.uol.edu.pk
[2]Department of Computer Science, University of Management and Technology
adnanabid7@gmail.com
[3]Department of Computer Science, University of Central Punjab
amjad.iqbal@ucp.edu.pk
[4]Department of Computer Science, University of Central Punjab
dr.farooq@ucp.edu.pk
[5]Department of Computer Science, Abdul Wali Khan University
shoaib.farooq@umt.edu.pk

ABSTRACT. *Higher education is delivered to the universities and recognized academics institutes. One of the biggest challenges to the higher education is to explore the data and use the data to improve the quality of education. In this paper we perform a case study on computer science courses of a university in Pakistan. This study is used to predict the grade of the student in a course on the basis of the previous grades. The association rule mining is used to find out the interesting pattern from the data. This technique is very useful for grade prediction. We performed the preprocessing of data to get quality results. In order to show the effective results, the data is divided into two parts, first part shows the preprocessed data and other was without preprocessed data. We compare the results and found reasonable accuracy.*
Keywords: educational data mining, association rule mining, course grade predictor

1. **Introduction.** Teacher evaluation plays an important role in predicting the grade of student . We can predict the grade of a student in a course if teacher evaluate the students correctly and honestly. This information helps in making decision to keep the quality of education high. Students are allocated sections on random basis as a result there is a mix quality of students in each section. Even if the instructor takes an average approach students at extreme will suffer. So it would be better if the there is less diversity in the class in terms of student quality. One can predict the grade in a course by passing his own parameters in the system. This way he would be in a better position to make decision if that is the right time to take the course or not. The students are future of the nation. The pass out students of university play a vital role in their field. If the students are passed out from the university in good grades it is better for their future progress. Sometimes the student takes more time to pass out their degree, because of improving grades in the course. The poor practice to pass the course is to enroll in the course until get best grade in the course. It will consume a lot of time and wastage of money as well. Most of the student cannot afford to get lower grade, to maintain or improve their

GPA. Another thing, which will make difference by these students, are the class environment. The teacher can perform well according to his/her abilities. When the student come to the class and has an idea of securing grade in that course, will eventually has positive effect on the attitude of the student.

In order to avoid these conditions, prediction of grade in the course before enrollment will be a good practice. The action happens and then human mental approach work on it. We can predict through previous record. To predict the grade in the course, it will be easy for student to take decision on enrollment of next courses. This will save the time and money as well. It will improve the quality of education. The enroll students are very determinant towards their goal in course either they will get more than the predicted grade. It is very good for the instructor to deliver their best with good environment in class. The prediction of grade can be accomplished through data mining technique [1-3].

One of data mining technique is association rule mining. It mines the interesting pattern from the data. The association rule mining can generate the strongest rule which shows the association of the course grades. One student get the good grade in C1 course what grade one can take in C2 course by finding the strongest rule of C1 and C2.

In this paper we are going to apply the association rule mining [4] on the grades of the student in a course. The data is divided into two parts. The data is preprocessed by cleaning the duplicate data, removing useless attribute, merging and grouping the data by finding maximum from the instance. The association rule mining is implement on both parts of the data. It will helpful to identify the quality of results.

2. **Paper Organization**. This paper has different section. Section 3 describe the motivation and contribution of this work. Section 4 give the literature review of the work. Section 5 defines the methodology of the work. Section 6 defines the result of our work.

3. **Contribution.** Our study will have significant effect on student as well as on educational environment. It will improve the student's decision making on course enrollment. The teachers have also dedicated and determinant students, which make healthy classroom environment. At the end, improve the quality of education. This study predicts the grade in the course through association rule mining.

4. **Related Work.** A lot of work is done to implement data mining techniques on the student record in semester to predict the performance of the student. The different algorithm of data mining has been used in education sector in different perspective. It is used to enhance the education system by discovering multiple patterns.Behrouz et al. [5] work on the prediction of final grade on the basis of homework done throughout the semester. There are tree classifier and non tree classifier applied on the web-based data of students.W.M.R. Tissera et al. [6] presents a real-world experiment conducted in an ICT educational institute in SriLanka. A series of data mining tasks are applied to find relationships between subjects in the undergraduate syllabi. This knowledge provides many insights into the syllabi of different educational programs and results in knowledge critical in decision making that directly affects the quality of the educational programs.Amelia et al. [7] present the G3P-MI to solve the problem of predicting a student's final grade based on his/her work in VLE from MIL perspective. The effectiveness of solving the problem is checked through representative paradigm of multiple instance learning and the results are compared. Experiments show that G3P-MI has better performance than the other techniques at an accuracy of 0.743 and achieves a trade-off between sensitivity and specificity at values of 0.702 and 0.775.Sajadin et al. [8] take University Malaysia Pahang (UMP) student data of semester 2007/2008 and describe the relationship between behavioral of student and their final academic performance. They apply decision tree j48 on the data and generate the strongest rule from it. The SSVM is applied on the rules to predicate student final grade. They apply k cluster to represent the strongest correlation between behavior of student and final grade. M.A Anwar et al. [9] applied Apiori algorithm on undergraduate engineering student English, mathematics and programming course data. The purpose is to provide guideline to education administration to improve and revise the teaching methodology and restructuring curriculum and modifying pre-requisite requirement of the various courses. This study identify that the student which perform better in English and mathematics will definitely perform better in programming courses. One student which is better in English only cannot perform better in programming course.Dr. Varun Kumar et al. [10] applied association rule mining on the graduate and post-graduate student. The study identifies the student's interest, curriculum design, teaching and assessment methodologies effect on student's grade. Another most interesting thing are found through this study, the post-graduate student which studied same course in graduate improved their grade.

5. **Methodology**. We have taken the computer science students grades data from one of the university of Lahore in Pakistan. Data has been taken from department to analyze and predict the grade. The data consist of 808 instances and 36 attributes. The type of data is nominal. There is an assumption of no inter dependency.All the attributes represent the courses offered in a period of time.

We have used WEKA for performing experiment. Excel sheet is used for preprocessing of the data.The possible values of the course are A+,A-,B+,B,C+,C,D+,D,F. It ranges from highest grade to lowest grade. The grade system followed by the university is A+,A-,B+,B,C+,C,D+,D,F. The data is divided into two parts. 1) preprocessed data 2) without pre-processed data. the association rule mining is applied on both the data.

5.1. **Preprocessing**. Data pre-processing is very important in knowledge discovery. The quality of analysis of result depends upon quality of data. The data gathering process are loosely controlled data The data can be noisy (containing errors or outlier values that deviate from the expected), incomplete (lacking attribute values or certain attributes of interest) and inconsistent (containing discrepancies). Data preprocessing consist of Data cleaning and Data transformation. Data cleaning include the cleaning of data by filling missing values, correcting noisy data, removing outlier and resolving inconsistencies. Data transformation operations like normalization and aggregation are additional data preprocessing procedures that would contribute toward the quality of mining interesting patterns.The student has enroll in a course multiple times to improve the grade in the course. We take the maximum grade from the repeated courses of student. The data is duplicated because of the course code are unique but the names are same. We merge those courses and get the proper grade. There are few steps involve in the pre-processing of data.

1. Grouping
2. Remove useless attribute
3. Remove useless instances
4. Merge
5. Convert into binominal

5.1.1. **Grouping**. There are different grades to the same course in data. The students took the course more than one time to improve the grade. There are multiple grades against one course of a student. The grades are grouped according to students grade in a course. The grouped grades are converted into number. We find the maximum number from the group. The numbers are converted into the grade.

Let S is the set of students {S1, S2,…Sm} take the courses {C1,C2,…..Cn}. The students took different grades in same course are grouped in set {g1,g2….gk}. Each element in grade set are values shown in table 2. In order to find out the maximum grade in a course. We have to take the following steps.

1. Replace the nominal values to numeric values shown in table 1.

2. Find out the maximum value from the set in equation 1

3. Replace the numeric value to nominal value.

$$Max(getgrade(SiCj)) \qquad (1)$$

Table 1. Numeric values given to grades

A+	9
A-	8
B	7
B+	6
C	5
C+	4
D	3
D+	2
F	1

5.1.2. **Remove Useless attribute**. The data consist of the attribute which have same value. It does not make difference in the result. We have to remove the useless attribute from the data.

5.1.3. **Remove Useless instances.** The data consist of useless instances. The useless instance in data is EX. There are 22 instances of EX we have remove such instances from the data.

5.1.4. **Merge.** The data consist of the different codes of the same course. We merge these courses. The maximum value are extracted from same courses with different codes. Let the course is C = {C1,C2, C3, ….Cn}, the course code is CC ={CC1,CC2,CC3…CCn}. If the same course has different course codes, e.g. CC1 and CC2,is the course code of the course C1. The maximum grade of the code of course is selected by using in equation2.

$$Merge(CCn(getmax(SiCj))) \qquad (2)$$

5.1.5. **Convert into Binominal**. The data consist of the attribute which have same value. It does not make difference in the result. We have to remove the useless attribute from the data.

6. **Mining Frequent items and Association**. Association rule mining finds interesting relations and/or correlation among large set of data items. In order to understand the mechanism of association rule mining on the data, it is discussed below.

Let { $C_1, C_2, … … ., C_{44}$ } be set of items of courses. Let D, be the set of database of students, where each student S is a set of items of grade such that S $\subseteq$ C. Each record shows the student grades in courses, this is represented as student identifier (SID). Let A be a set of grades in courses. A student has grade in course A if and only if A $\subseteq$ S. The implication form of association rule is A$\Rightarrow$B, where A $\subset$ S, B $\subset$ S, and A $\cap$ B $= \varnothing$

Support (s) and confidence (c) are the parameter to measure the rule interestingness. They reflect the effectiveness and confidence of the discovered rule respectively. A support of 4% of the rule A$\Rightarrow$B means that A and B exist together in 4% of the student grade database. The rule A$\Rightarrow$B having confidence of 60% in the student grade database means that 60% is the percentage of student grade database in D containing A that also contains B.

7. **Result and Discussion.** We have used Apriori to get the interesting rules because we had not the large amount of data however if that would be the case frequent pattern growth could be applied to overcome the performance barrier. We have experimented with different values of min support and confidence to get the interesting rules. We have found some interesting hidden rules that could be of significant help for the academician in decision making. The top 5 rules are mentioned in table 2. The Rules gives an insight in the system that is also validated by Teachers e.g Rule 1 Suggests that those students who are good in Mathematics gets good grades in programming courses that are offered earlier in degree. Rule 2 Suggests that those who are not good in the prerequisite courses maintains there grades in the subsequent course. Rule 3 and Rule 4 Suggest .Those who are good in the humanities courses are unlikely to get good grades in programming courses.

These rules will be helpful for the new comers enrolled in the earlier semester. They will be familiar how the core courses will help them to maintain the CGPA. These result will help student to decide the enrollment in the course.

In future we will also consider the other physiological and environmental behavior as parametrs. We will infere the rules by using intelligent techniques. We will optimize the rules to predict the grades in better way.

Table 2. Top-5 rules from Weka

Rule	Detail
1	calculus=A -> computer programming=A
2	PF = C -> OOP = C
3	English=A and communication skill=A -> computer programming=C
4	Programming=A and calculus=A and oop=A-> management=C
5	LinearAlgebra = A -> Computer Vision = B

REFERENCES

[1] Bidgoli, B.M., Punch, W.F. (2003).Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System. *Genetic and Evolutionary Computation*. (Vol. 2, pp. 2252–2263)..

[2] Superby, J.F., Vandamme, J.P. (2006). Determination of Factors Influencing the Achievement of the First-year University Students using Data Mining Methods. *EDM'06: Workshop on Educational Data Mining*. (pp. 37-44)..

[3] Kotsiantis, S.B., Pintelas, P.E. (2005). Predicting Students Marks in Hellenic OpenUniversity. *ICALT'05: The 5th International Conference on Advanced Learning Technologies*. (pp. 664-668)..

[4] Agrawal, R., Imielinski, T., and Swami, A.(1993). Minning association rules between sets of items in large database. *ACM SIGMOD international conference on management of data*,(Vol. 22, pp. 207-216)..

[5] Bidgoli, B. M., Kashi, D. A., Kortmeyer, G. and Punch, W. F.(2003). Predicting student performance:an application of data mining method with an educational web-based system. *IEEE international conference on Frontiers in Education*,(Vol. 1, pp. 13-18)..

[6] Tissera, W. M. R., Athauda, R. I., and Fernando, H. C. (2006). Discovery of strongly related subjects in the Undergraduate Syllabi using Data Mining, *IEEE International Conference on Information Acquisition*, (pp. 57-62)..

[7] Zafra, A., Ventura, S. (2009). Predicting student grades in learning managemnt systems with multiple instance genetic programming. *2nd international conference on educational data mining*. (pp. 309-318)..

[8] Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., and Wani, E., Prediction of academic performance by an application of data mining technique. *international conference on management and artificial intelligence*. (Vol. 6)..

[9] Anwar, M. A., Ahmed, N., and Khan, W. (2012). Analysis of students' grades in Mathematics English, and Programming courses:A KDD approach. *International   journal of future computer and communication*,(Vol. 1, no. 2, pp. 111-115)..

[10] Kumar, V., Chadha, A.(2012). Mining association rules in Student's assessment data. *IJCSI*. (Vol. 9, pp. 211)..