International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

PERFORMANCE EVALUATION OF STUDENT BASED ON ATTENDANCE USING DATA MINING TECHNIQUE

Mrs. Kirti Kakde[1], Dr. Praveen Gupta[2]

[1]*Assistant Professor, Dept. of MCA,YMT College of Management, Kharghar-4102010,Navi Mumbai, Maharashtra*

ABSTRACT - *Prediction of students' performance becomes more challenging due to the tremendous volume of data in educational databases. The patterns are hidden in the academic data set and it is identified through various data mining algorithms. Student performance can be predicted from the data available in the institution. As large data is available which helps to calculate student expected grade in the exam. In this paper data mining techniques are explored to evaluate student's performance based on attendance using Association Rule Mining. A priori algorithm is used for finding associations among various attributes. Based on assignments, class test grade, lab work, last semester grades and attendance student performance is evaluated. This helps to predict whether the students performance affects based on the special factor attendance*

***KEY WORDS*: Association Rule Mining, Data mining Technique, A priori Algorithm, Class attendance, Student performance, Grade.**

1. INTRODUCTION

Educational institutes are offering number of courses to the students, where class attendance is an important factor along with other factors. But as the institute is having huge data which is generated after every exam, this data is difficult to evaluate for student performance. This extracted data need proper method of extracting information for assessment. Various activities students are performing during their academics, based on which prediction can be done such as grade. From the result we can analyze that whether the institution should change their attendance policy. In many institutions based on the attendance the student is not allowed for the exam due to which a good student having poor attendance is not able to give exam. To predict student performance variables such as last semester grades, class test grades assignment, lab work and attendance are taken into consideration. Based on this, the data of some students is mined for pattern and prediction of students' last semester grade affects based on the attendance. Institutions are monitoring student attendance which is linked to the performance. Other factors are also taken into account which is finally compared with attendance. In this paper, support and confidence is calculated with attendance. Support and confidence is also calculated with other factors such as class test grade, assignments, lab work and last semester grade.

In this paper data mining methodologies has been used to predict student performance based on the attendance. This paper investigates the accuracy using Association rules for predicting student performance. Using these generated mining rules, students are identified and monitored by the faculties which ultimately help to enhance their academic grades.

2. RESEARCH METHODOLOGY

In this paper, Weka tool is applied on the students dataset. For analysis purpose the dataset file in the form of .arff is used. It is also well-suited for developing new machine learning schemes. This tool provides all support and confidence for mining algorithm. Apriori algorithm with minimum support and confidence value with lift value can easily used to achieve mining results. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset.

2.1 APRIORI ALGORITHM

The Apriori Algorithm is used for mining frequent itemsets for association rules. This algorithm may be considered to consist of two parts. In the first part, those itemsets that exceed the minium support requirement are found. Such item sets are called frequent itemsets. In the second part, the association rules that meet the minimum confidence requirement are found from the frequent itemsets.[1]

2.2. ASSOCIATION RULE MINING

Association rule learning is a prominent and a well-explored method for determining relations among variables in large databases.

Let $I = \{i_1, i_2, i_3, \ldots, i_n\}$ be a set of n attributes called items and $D = \{t_1, t_2, \ldots, t_n\}$ be the set of transactions. It is called database. Every transaction, t_i in D has a unique transaction ID, and it consists of a subset of itemsets in I. A rule can be defined as an implication, $X \longrightarrow Y$ where X and Y are subsets of $I(X, Y \subseteq I)$, and they have no element in common, i.e., $X \cap Y$ [2].

Support indicates the frequency of the pattern. A minimum support is necessary if an association is going to be of some business value.

- $$supp(X) = \frac{\text{Number of transaction in which } X}{\text{Total number of transactio}}$$
  .

- **Confidence** of a rule is defined as follows:

$$conf(X \longrightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

3. DATA COLLECTION AND PRE-PROCESSING

In this research paper, data is acquired from YMT College of Management, affiliated to University of Mumbai. Before applying the Data Mining techniques the collected raw data needs to be pre-processed into an expedient format. On the basis of collected data, some attributes are considered to evaluate students 'performance. The Attributes used for classifying the student's performance are mentioned in Table I.

Table -1: Attributes and Description

Variables	Description	Possible Values
Last Semester Grade	Grade obtained in Previous Semester	{O>=80,A>=75,B>=70,c>=60,d>=55, E>=50,P>=45,F<45}
Class Test grade	Grade based on marks obtained	{Fail, Average, Good, Excellent, Outstanding}
Assignment	Assignments submitted	{Yes, No}
LW	Lab Work	{Yes, No}
Atten	Attendance	{Good, Average, Poor}

Table II. SAMPLE of Student Data Collection

Student Roll No	LSM	Class Test grade	Assignment	LW	Attendance
1	**A**	**Good**	**Yes**	**Yes**	**Good**
2	A	Good	Yes	Yes	Average
3	A	Average	Yes	Yes	Poor
4	B	Average	Yes	No	Poor
5	A	Good	Yes	Yes	Good
6	C	Average	No	No	Poor
7	A	Good	Yes	Yes	Poor
8	A	Good	Yes	Yes	Average
9	A	Good	Yes	Yes	Poor
10	B	Good	Yes	Yes	Average
11	C	Average	No	No	Average
12	A	Good	Yes	Yes	Average
13	A	Average	Yes	Yes	Poor
14	A	Good	Yes	Yes	Average
15	A	Good	Yes	Yes	Average

4. SAMPLE OF RULES GENERATED

Support of student performance with A grade and attendance:

Rule No.	Rules	Support	Confidence
1	A, Good	0.13	1
2	A, Average	0.33	1
3	A, Poor	0.26	1

Support and Confidence of last semester grade=good,Class test grade=good, assignment=yes, lab work=yes and attendance(good, average and Poor) to Class test grade, assignment, lab work

Rules Generated:

Confidance{lsm,ctg,assi,lw,atten}->{ctg,ass,lw}

Student Roll No	LSM	Class Test grade	Assignment	LW	Attendance	Support	Confidence
1	A	Good	Yes	Yes	Good	0.13	0.2
2	A	Good	Yes	Yes	Average	0.33	0.5
3	A	Average	Yes	Yes	Poor	0.13	0.5

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: attendance
Instances: 15
Attributes: 5
lsm
ctg
assi
lw
attendance
=== Associator model (full training set) ===
Apriori
=======

Minimum support: 0.7 (10 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 6

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 5

Size of set of large itemsets L(3): 2

Best rules found:

1. lw=yes 12 ==> assi=yes 12 conf:(1)
2. lsm=a 10 ==> assi=yes 10 conf:(1)
3. lsm=a 10 ==> lw=yes 10 conf:(1)
4. ctg=good 10 ==> assi=yes 10 conf:(1)
5. ctg=good 10 ==> lw=yes 10 conf:(1)
6. lsm=a lw=yes 10 ==> assi=yes 10 conf:(1)
7. lsm=a assi=yes 10 ==> lw=yes 10 conf:(1)
8. lsm=a 10 ==> assi=yes lw=yes 10 conf:(1)
9. ctg=good lw=yes 10 ==> assi=yes 10 conf:(1)
10. ctg=good assi=yes 10 ==> lw=yes 10 conf:(1)

5. RESULTS AND DISCUSSIONS

The above results discover apriori algorithm is used to obtain minimal rules. From the extracted pattern apriori algorithm is used in predicting student last semester grades depend on the variables. From the result it is found that the impact of class attendance on the result is very low. From the above results, it is found that just 13% of the students has good attendance and had "A" grade, 33% of the students had average attendance and achieved A grade and 26% of the students had poor attendance and they also achieved A grade. 13

6. CONCLUSION:

Student performance is predicted using data mining rules. Using these rules it is predicted that the impact of student attendance on student performance is very low. A Student with good attendance can also fall into any grade category from "A-F". This indicates that class attendance is not actually the factor that determines student academic performance but other key factors such as class test marks, assignment, lab work are equally important.

REFERENCES

[1]. Asif, M. and J. Ahmad (2015) Data Analysis of Effectiveness of Apriori and Frequent Pattern Tree Algorithm in Software Engineering Data Mining International Conference on Intelligent Systems, Modelling and Simulation, Vol. 1, IEEE.

[2]Abu, T. and M. El-Halees (2012) Mining Educational Data to Improve Students' Performance: A Case Study International Journal of Information and Communication Technology Research, vol. 2, no. 2.

[3]Margaret H. Dunham, "Data Mining: Introductory and advanced Topics", Pearson 2013.

[4]Romero, C, Ventura, S. and Garcia, E. (2008) Data mining in course management systems: Moodle case study and tutorial Department of Computer Sciences and Numerical Analisys, University of Córdoba, 14071 Córdoba, Spain

[5] Vipin Kumar, Pang-Ning Tan, Michael Steinbach, "Introduction to Data Mining", Addison-Wesley, 2005. ISBN : 0321321367