Analysis: Classification Data Mining Process in Primary Education System

Dr. Manmohan Singh
Asst. Prof., Department of Computer Science & Engineering, ITM Universe, Vadodara, India
Manmohan_sati@yahoo.co.in

Dr. Amit Dutta
Deputy Director, All India Council for Technical Education, Vasant kunj, Nelson Mandela Marg, New Delhi, India
amitdutta07@gmail.com

Dr. Ramesh Prasad Aharwal
Asst. Prof., Department of Mathematics
Govt. P.G. College, Damoh (M.P.)
ramesh_ahirwal_neetu@yahoo.com

Abstract: In this paper, we will discuss about Educational system. Data mining is useful in education system. Using Data mining techniques to improve the efficiency of educational institutions based on Based on Student class Attendance one of the most useful data mining techniques for primary education is classification. Classification maps data into predefined group of primary classes. Classification is a supervised learning approach because the classes are determined before examining the data. The prediction of student’s performance with high accuracy is more beneficial for identifying low academic performance of the students at the beginning. Classification is the processing of finding a set of models which describe and distinguish data primary classes or concepts. The derived results may be represented in various forms, such as classification (IF-THEN) rules, decision trees, or neural networks. Models then can be used for predicting the class label of data objects.

Keywords- Data Mining, Educational Data Mining, Zero R Classification, Naïve Bayes Algorithm, Decision Tree.

1. INTRODUCTION:

A Data Mining Process can be defined and applied for exploring and analysis data to identify useful pattern. Data Mining is a process of extracting previously unknown, valid, potential useful and hidden patterns from large data sets [1]. Data Mining task can be classified into two categories: Descriptive and Predictive. Descriptive mining tasks characterize properties of the data in a target data set. Predictive mining tasks perform induction on the current data in order to make predictions [3]. Predictive mining
task is clustering, prediction and descriptive mining task is association rule and summarization. Provide the Quality education to student is the main goal of education system. For the improvement and development in education system, data mining can be very useful. Data mining techniques can help for future prediction about student performance. Data Mining can help for partition students into homogeneous groups according to their characteristics and abilities [2] nowadays; educational organizations are face is a very high competitive environment and everyone wants becomes first position for teaching. Data mining provides many tasks that could be used to study the student performance.

2. RELATED WORKS:

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. C.Romero in 2008 [1] Gave case study of using educational data mining in Moodle course management system. They have described how different data mining techniques can be used in order to improve the course and the student’s learning. All these techniques can be applied separately in a same system or together in a hybrid system. Han j. and Kamber 2006 [3] in Explained that k-means is a well known clustering algorithm tends to uncover relations among variables already presented in dataset. Kifaya, in 2009 [5] Explained that K-means clustering is a widely used method that is easy and Quite simple to understand. Cluster analysis describes the similarity between different cases by calculating the distance. These cases are divided into different clusters due to their similarity. Shaeela Ayesha in 2010 [7] Have a described the use of k-means clustering algorithm to predict student’s learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students. Henrik in 2001 [8] Concluded that clustering was effective in finding hidden relationships and associations between different categories of students.

3. DATA MINING IN EDUCATION:

In the education field, Educational Data Mining is application of Data mining. On Process transforming raw data compiled by education system is called educational data mining. Extracted information is useful for management decision making process, teachers as well as students [4]. With the help of data mining in education, we can classify students in to groups. One group is need more guidance means have poor performance and other have less performance means need less guidance. In educational data mining following steps are performed.
1. **Zero R Classification Algorithms:**

The most primitive learning scheme in Weka Zero R Classification predicts the majority class in the training data for problems with a categorical class value, and the average class value for numeric prediction problems. It is useful for generating a baseline performance that other learning schemes are compared to. In some cases, it is possible that other learning schemes perform worse than Zero R, an indicator of substantial over fitting.

2. **Decision Tree Based on Student Attendance:**

This section presents the class attributes details and which parameters have been taken in to considerations during creating a decision tree model. Class attribute consists four of classes as shown in figure and parameter setting is shown in figure Generated decision tree with J48 classifier attendance.

3. **Naïve Bayes Algorithm:**

Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers Algorithm applies on the Grade previous result it indicates about the true positive, false negative, true negative, false positive.

4. **BENEFITS IN EDUCATION SYSTEM USING DATA MINING:**

(1) Improving the student’s performance based on extracting knowledge.
(2) analyzed student academic data and enhance the quality of education system.
(3) Improving student’s learning process.
(4) Improve research capability in the educational data.
(5) Time saving during knowledge extraction for educational data.
(6) More effective sharing of information in educational area.
(7) Data mining techniques can be utilized effectively in selecting course, managing student’s improving attendance / dropouts providing supplementary classes where necessary.
(8) Data mining is useful for analyze student’s data for predict their learning behavior and to warn students at risk before their final exams.
5. RESULTS ANALYSIS

1. Zero R Classification Based on Student Attendance:

Urban area Government school attendance number of irregular students is 10 out of which boy -4 and girls -6. Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers Algorithm applies on the Grade previous result it indicates about the true positive, false negative, true negative, false positive.

2. Naïve Bayes Algorithm Based on Student Attendance:

In Rural area government school attendance, number of irregular students out students is 60 with Boys – 41 and Girls -19. In Rural area private school attendance, number of irregular student out is 70 of which boys -34 and girls -36.
In Urban area private school attendance, number of irregular students is 10 students out of which boys -2 and girls -1. Urban area Government school attendance number of irregular students, out of 10 is students are 9 out of which boy -4 and girls -5. All the attendance of primary school most of government school student attendance is low as compare other school attendance so attendance is directly affected result of students.

3. Decision Tree Based on Student Attendance:

This section presents the class attributes details and which parameters have been taken in to considerations during creating a decision tree model. Class attribute consists four of classes as shown in figure and parameter setting is shown in figure Generated decision tree with J48 classifier attendance.

![Figure 3 Applying Decision Tree with J48 Classifier Based On Student Attendance](image)

4. Classification Accuracy:

Note that the previously presented classification algorithms are neither Student school Attendance, student school Location and Previous Result in Grade. For example, someone might object to your use of the simple minimum distance to means classifier by pointing out that your school classes have different variability, or a student might object to your use of a Maximum Likelihood Classifier because our data is not normally distributed. These algorithms are neither regular” &“Irregular “but rather they do a Regular or Irregular in classifying. Accuracy and Cost Efficiency assessment tries to base on attendance, School type, and pervious result in grade how good any one algorithm was done by the classifier.
5. Accuracy Based on Student Attendance:

Table 1 Performance Compression Approach Based on Student Attendance- Naïve Bayes, Decision Tree and Zero R (Accuracy)

<table>
<thead>
<tr>
<th></th>
<th>Naive Bayes</th>
<th>Decision tree</th>
<th>Zero R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>6.6</td>
<td>58.209</td>
<td>58.209</td>
</tr>
<tr>
<td>Irregular</td>
<td>94.243</td>
<td>76.119</td>
<td>76.119</td>
</tr>
</tbody>
</table>

Figure 4 Performance Compression Approach Based on Student Attendance- Naïve Bayes, Decision Tree and Zero R (Accuracy)

6. CONCLUSION:

In this paper we have discussed the based on student attendance data mining in education system and data mining techniques such as clustering and classification now a Zero R classifier is a simple classifier. However although it is simple, Zero R can outperform more sophisticated classification methods in Classification accuracy, based on Attendance. Besides that it has also exhibited high accuracy and speed when applied to large database. Moreover, it is very fast for both learning and predicting. Its learning time is linear in the number of examples and its prediction time is independent of the number of examples. Zero R classifier is also fast, consistent, easy to maintain and accurate in the classification of attribute data and from computation point of view, Zero R is more efficient. Application of classification data mining approach is based on this criteria based on School type (Simulation), location Based school and students results Previous Grade History. Following classification techniques are used.
References


