Learning Prediction Using Machine Learning Techniques for Students

Semon Raj .A  
M.Phil (Research Scholar)  
Islamiah College (Autonomous)  
Vaniyambadi

Dr.H.Faheem Ahmed., M.Tech., Ph.D  
Head of the Department of Computer Science  
Islamiah College (Autonomous)  
Vaniyambadi

Abstract: Now a day’s e-learning is smartly growing technology. This technology is more helpful for students to communicate with their professors through chats or emails. E-Learning also removes the obstacle of physical presence of an Elearner. The main aim of this paper is to predict student performance in their final exams using different machine learning techniques. Information like attendance, marks, assignments, class participation, seminar, CA, projects and semester are collected to predict student performance. This prediction helps the instructors to analyze their students based on their performance. For that we have used WEKA tool for the prediction of the student performance. WEKA (Waikato Environment for Knowledge Analysis) is one of the data mining tool which is used for the classification and clustering using data mining algorithms. This prediction helps the students and the staffs to know how much effort their students need to be put in their final exams to get good marks.

Keywords- E-learning, Data Mining, Machine Learning algorithms, Data Pre-processing.

I. INTRODUCTION

The main aim of educational organizations is to provide better and excellence education to their students. The best way to achieve better quality in education is by determining the understanding of students in a particular course. At present, in educational institutions, data mining plays a vital role and it is the most essential area of research with the motive of finding significant data from knowledge keep in huge dataset. Educational Data Mining (EDM) could be a main analysis space to predict helpful data from educational information to boost the standard of educational performance, to know higher and to own better assessments of students’ learning. EDM may be taken mutually of the training sciences and as a field of information mining.

EDM can be applied in modelling the knowledge, behavior and experience of the user. Data Mining or knowledge discovery has emerged as an important area to analyze data from different perspectives and also helps to summarize it into useful information. Educational Data Mining uses several techniques like call trees, neural networks, naive Bayes, support vector machines and lots of others. WEKA is one of the open source software intended for data analysis and knowledge discovering. It implements a wide range of machine learning algorithms and is widely used in data mining applications. The main aim of this paper is to predict student’s performance using data mining algorithms and to find the highest accuracy between the classifiers.

The paper is ordered as follows. The second part of the paper represents literature review and similar research. The third part presents the datasets and description. The fourth section represents the data preprocessing, experiments and discussions. Fifth section of the paper presents conclusions and scope for the future research. The references are presented in the last section of the paper.

ILITERATURE REVIEW

The researchers have referred various papers that uses different methods and methodologies to analyze and predict student results.

1. E-learning student assessment by means of evaluation strategy [1]: In this research article, they have assessed and understood the student performance using metacognitive technique. The study comprises the assembling of model and explaining mechanism. In this paper they have proposed Online Learning Stratagem. The area of perceived-skill line of attack, particularly the understanding strategies is one of the chief drawbacks of this research.
2. Analyzing student performance using data mining techniques [2]. In this paper, they have used number of parameters to predict the student performance based on their academic performance. The authors have used different assessment techniques to evaluate the student performance. The focus of this research is to use hidden Markov model to grades the students, which has used totally different learning power and substitute assessment technique. Consistent with the experimental analysis, the new evaluating strategy is considerably useful in evaluating performance of scholars.

3. Evaluation of student’s performance by means of Data Mining Techniques [3], the analysis proposes technique accustomed assess the coed performance mistreatment varied prediction indicators. The most focus is to predict the students’ performance. The values obtained during this analysis is showed with the intention of information mining technique available trained analyze tools for learner knowledgeable performance with 94 % success analysis.

4. Early prediction of student performance using machine learning algorithms[4]: in this paper, they have used some undergraduate college students’ details to predict the performance. The main objective of this paper is to predict weak and strong students, so that some form of remediation may be organized for them. And the future selection algorithms were applied on the datasets to reduce the number of features. Five classes of machine learning algorithm were trained and then tested.

5. Machine learning approach to predict student academic performance [5]: in the education research study various aspects like social, economic, personal, cultural, geographical, institute environment and others are evaluated. Such aspects may either help a student to shine during the academic period or halt academic program. Such failure is known as drop-out and data mining algorithm helps in finding the factors that mostly contribute to the student’s performance. In this research machine learning approach or machine learning classification models are to be used, to predict students’ academic performance. This is fit for Educational data mining. In this research study, the students’ performance are predicted and the results of these prediction models are compared. This research also discusses how these machine learning models can help to improve an education system by considering the different factors in terms of accuracy, specificity, precision, prevalence, recall, F-Measure etc., in results. In addition to that, this research is carried on by applying Standard and Hybrid approaches as well as fuzzy logic and compared outcomes. This research also helps in predicting the performance of other state students by identifying the appropriate factors. The main aim of this research is to predict the performance of students and to identify the most appropriate model for that; so that on the basis of that we would be able to take the appropriate steps to improve their performance and to cut down the dropout rate.

III. DATASETS AND DESCRIPTION

Student Data Set consist of 1100 records of students with their Attendance, marks, assignments, class participation, seminar, CA, projects and semester. The datasets were split into two parts one is trained data and the test data, trained data consist of 1000 data’s and the test data consist of 100 data in a datasets. The following table shows attribute with their descriptions and values.

<table>
<thead>
<tr>
<th>ATTRIBUTE</th>
<th>DESCRIPTION</th>
<th>VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>Student attendance</td>
<td>P&gt;60%, F&lt;60%</td>
</tr>
<tr>
<td>Assignment</td>
<td>Assignment</td>
<td>P&gt;=3, F&lt;3</td>
</tr>
<tr>
<td>Class Participation</td>
<td>Class participation of the</td>
<td>P&gt;=3, F&lt;3</td>
</tr>
<tr>
<td></td>
<td>students in the class.</td>
<td></td>
</tr>
<tr>
<td>Seminar</td>
<td>Performance of the seminar.</td>
<td>P&gt;=3, F&lt;3</td>
</tr>
<tr>
<td>CA</td>
<td>Internal exam marks.</td>
<td>P&gt;=3, F&lt;3</td>
</tr>
<tr>
<td>Project</td>
<td>Project</td>
<td>P&gt;=3, F&lt;3</td>
</tr>
<tr>
<td>Semester</td>
<td>Semester marks</td>
<td>P&gt;=35, F&lt;35</td>
</tr>
</tbody>
</table>
IV DATA PREPROCESSING
WEKA is a data processing tool developed by Department of engineering, University of Waikato, New Sjaelland. Its associate degree open supply code provide out below the overall Public License (GNU). Rail is cluster of machine learning algorithms for the data mining tasks.

These algorithms is applied on to the info or referred to as from the Java code.

The pre-processing is an important step that is used to extract and improve the quality of data. WEKA tool import dataset from a proper file like attribute relation file format which is the preferable one. The below figures shows that the output of data pre-processor and data model visualization, respectively.

Fig. 1. Screenshot of trained dataset

Fig. 2. Screenshot of Test dataset

Fig. 3. Model Visualization
EXPERIMENTAL RESULTS AND DISCUSSION

A comparative analysis has been done on set of specified dataset using some machine learning algorithms. Those algorithms are BayesNet, J48, Random Forest, Navie Bayes, Random Tree, Decision Stump, Hoeffding Tree and LMT. Each algorithm has its own characteristics to classify the dataset. Following table shows the predicted results of all classifiers.

Table-II: classifiers with the results

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Classifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BayesNet</td>
</tr>
<tr>
<td>Accuracy %</td>
<td>100</td>
</tr>
<tr>
<td>Correctly Classified Instance</td>
<td>100</td>
</tr>
<tr>
<td>Incorrectly Classified Instance</td>
<td>0</td>
</tr>
</tbody>
</table>

In the above table shows that the Bayesian Network, J48, Random Forest, REPTree and LMT algorithms has more correctly classified instance compared to other algorithms-
The below two graphs shows the analysis of all classifiers based on the RMSE and RRSE. It shows that the J48 and REPTree has the minimum error values among other classifiers.

Figure 4: ACCURACY OF THE CLASSIFIERS

Figure 5: GRAPH OF ROOT MEAN SQUARE ERROR VALUE

Figure 6: GRAPH OF ROOT relative SQUARE ERROR VALUE
The above chart shows that the performance accuracy of the five classifiers based on different classification metrics. These metrics are: Precision, Recall and F-measure measure are very important to determine the classifiers based on the accuracy. These metrics shows that Bayesian Network classifier performs better than other classifiers.

**COMBINATION OF TWO ALGORITHMS USED FOR THE PREDICTION**

Here I have combined two algorithms for the prediction of student’s performance. And the below table shows that the J48 with REPTree, BayesNet with LMT and LMT with Hoeffding Tree produced best result compared to others and the result shows that these combination of algorithms are the best classifiers.

**TABLE 3: combined classifiers with results**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Classifiers</th>
<th>BayesNet</th>
<th>J48 with REPTree</th>
<th>Random Forest with Random Tree</th>
<th>J48 with Random Forest</th>
<th>BayesNet with LMT</th>
<th>LMT with Hoeffding Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy %</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Correctly Classified Instance</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Incorrectly Classified Instance</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
The above graph shows the accuracy of the combined classification algorithms. And the below two graphs show the analysis of all classifiers based on the RMSE and RRSE. It shows that the Random Forest with Random Tree combination has the minimum error values among other classifiers.

**FIGURE 7:** GRAPH OF ACCURACY OF COMBINED ALGORITHMS

**FIGURE 8:** GRAPH OF ROOT MEAN SQUARE ERROR OF COMBINED ALGORITHMS
CONCLUSION

Student evaluation is one of the important tasks in instructional data processing. This prediction discovers the likelihood to forecast the success rate of the scholars within the specific course using machine learning algorithms. From the above discussion, we can conclude that Bayesian Network, J48, Random Forest, REPTree and LMT algorithms for predictive modelling gave us the best result as compared to the Decision Stump, Naive Bayes, and Hoeffding Tree using single algorithms. And also we have used combination of two algorithms to predict student academic performances, in that we have got better results. J48 with REPTree, Random Forest with Random Tree, J48 with Random Forest, and BayesNet with LMT these combination of the classifiers gives 100% result. With this paper, we try to find out the best result with only selected attributes. This paper helps other researchers to find out which are the algorithms are best to predict student performance. In our future work, we try to implement our work on the different dataset with different classification algorithm. By doing this we can able to find the best algorithm which is going to perform almost perfect on every dataset in hand.

AUTHORS PROFILE:

Semon Raj A has completed his M.C.A and is pursuing his M.Phil in Computer Science. His research area includes Machine Learning and Block chain technologies. His passion is to further his research in Block chain technology and contribute to the advancement of humanity.
REFERENCES

7. C. Vera, A.Cano, C. Romero, S. Ventura, “Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data”, DOI 10.1007/s10489-0120374-8