Efficient Data Mining Techniques to Improve Academic Performance of Students

Rashmi V. Varade¹, Blessy Thankanchan²
¹,² Jaipur National University, Jaipur, India; *e-mail: rashmivvarde@gmail.com

ABSTRACT
Predicting the academic performance of students is very challenging due to large volume of data in the educational institutions database. Data mining techniques are implemented to predict students’ academic performance in many institutions. Because of predicting students’ performance, it will help teachers and institutions to decide strategies to teach to the students who are weak in studies and also they can define different strategies who are good in studies so that these students can perform better.

So, aim of this paper is to study such a data mining technique which will help us to predict students’ academic performance in advance.

Key Words: Data mining. Decision tree, Naïve bias, KNN.

INTRODUCTION
In the recent years there has been increasing interest of applying data mining algorithms in all fields such as medicine, education, business, marketing, engineering for large volume of data analysis. Data mining is process of turning the raw data into meaningful and useful data. It is also called as Knowledge discovery in database. It includes data cleaning, data integration, data transformation, data mining, knowledge presentation and pattern evaluation. So, it is very essential to use data mining techniques in education also.

But as we know there is no fix pattern to evaluate academic performance of student in India.

Most of the institution measure student’s performance on grades given to them and internal assessments and projects. Practical examination and viva voce.

Data mining is a powerful artificial intelligence (AI) tool, which can discover useful information by analysing data from many angles or dimensions, categorize that information, and summarizethere relationships identified in the database.[¹]
SEARCH STRATEGY FOR LITERATURE REVIEW

For this research various research articles, journals, conference papers from IEEE, ResearchGate, Elsevier, Springer, ACM from 2010 to 2020 has been referred. And, workshop documents, blogs web sites lot of PDFs and books about Data mining, Data mining techniques are referred.

Important Factors of Students Used for Predicting Student’s Performance:

There are several factors like economic, social, geographical, Psychological effects on the academic performance of the students. From past few years several researchers have working on it so we reviewed 25-30 research papers to find out factors affecting on students’ academic performance. Omar Augusto Echegaray-Calderon and Dennis Barrios-Aranibar in his research paper showed that age, parents’ educational qualification, parents' occupation, hard work, math grade, science grade, reading grade effects on academic performance.[2]

Brijesh Kumar Baradwaj and Saurabh Pal in their research paper stated that Previous Semester Marks, CTG Class Test Grade, SEM Seminar Performance, ASS Assignment

GP General Proficiency, ATTAttendance, LW Lab Work End Semester Marks effects on students’ academic performance.[3] So most important factors that affects on academic performance are grades, age, gender, marital status, admission type, study hours.[4-8]

Different Data Mining Techniques used for Predicting Student’s Performance

In data mining there are various algorithms used for prediction of students’ academic performance some are as follows

Decision Trees, KNN, Naïve Bias, Random Forest.

Decision Tree

A decision tree is widely used mining technique much because of the fact that the output is very easy and intuitive to interpret. It is flow-chart-like tree structure where all internal nodes have two or more child nodes. The leaf nodes denote the decision made or the class label, the arcs the condition that we have applied and the internal nodes denote the attributes. Decision trees are amongst the favourite among researchers applying the data mining techniques on the educational domain.[2]

KNN (K-Nearest Neighbour)

K-Nearest neighbour classifier represents a totally dissimilar approach to classification. They do not build any clear universal model, but estimated it only locally and implicitly. The main idea is to classify a new object by examining the class values of the K most alike data points. The selected class can be either the most frequent class among the neighbours or class distribution in neighbourhood.[3]

Naive Bayes

Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.

Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting [4]. There are so many algorithms available for prediction in data mining but in this paper, we only considered decision tree, KNN and Naïve Bayes algorithms. The table -1 shows brief review of research papers with name of the author, algorithm applied and its accuracy is shown.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Decision tree</th>
<th>KNN</th>
<th>Naïve Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nguyen Hua Nghe, Paul Janecek, and Peter Haddawy</td>
<td>84.18%</td>
<td>--</td>
<td>78.57%</td>
</tr>
<tr>
<td>Biza Hasm, Sellaplan Palamavam, Abdul Raffi and Abdul Raffu</td>
<td>63.63%</td>
<td>--</td>
<td>90.90%</td>
</tr>
<tr>
<td>C. Anuradha and T. Veluswaran</td>
<td>--</td>
<td>68.32%</td>
<td>72.51%</td>
</tr>
<tr>
<td>P.V. Pravnen Sundar</td>
<td>--</td>
<td>--</td>
<td>56.1%</td>
</tr>
<tr>
<td>S. Ramesh, P. Parkavi, K. Ramar</td>
<td>--</td>
<td>--</td>
<td>49.5%</td>
</tr>
<tr>
<td>Yasmeen Shaker Alshain, Nancy Khamees Ab Halim, Eman Saleh AlNagi</td>
<td>66%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>M. Shabir Ali Haqab, Alland Tucker and Leela Yousef</td>
<td>76.70%</td>
<td>--</td>
<td>73%</td>
</tr>
<tr>
<td>Mahesh A. et. al.</td>
<td>--</td>
<td>--</td>
<td>91%</td>
</tr>
<tr>
<td>Dr. S. SENTHAL, WONG MU LIN</td>
<td>--</td>
<td>--</td>
<td>88.44%</td>
</tr>
<tr>
<td>Matchi Kumar, Yass Kelliyath Sehal</td>
<td>98.86%</td>
<td>83.00%</td>
<td>98.00%</td>
</tr>
<tr>
<td>Najmus Saher Shab</td>
<td>84%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>R. Asif et. al.</td>
<td>83.6</td>
<td>74%</td>
<td>73%</td>
</tr>
<tr>
<td>M. Koutina et. al.</td>
<td>68.5%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>N. Giga et. al.</td>
<td>99.9%</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>N. V. Shinde et. al</td>
<td>91.5%</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table-1: shows data mining techniques applied for student’s academic performance
**DISCUSSION**

In this section, we will discuss the findings of our analysis. After reviewing several research papers, we have found that the most commonly used algorithms to predict students' academic performance are Decision tree, KNN, and Naïve Bayes. From the above research papers, we found that the maximum accuracy of predicting student's academic performance using Decision tree algorithm is 99.9% while the minimum accuracy is 63.63% and the average accuracy is 82.28%. Where as in case of KNN algorithm, the maximum accuracy of predicting students' academic performance is 100% and the minimum accuracy is 68.32% and the average accuracy is 81.33%. And Naïve bayes algorithms' maximum accuracy of predicting students' academic performance is 100% and the minimum accuracy is 49.5% and the average accuracy is 79.20%.

<table>
<thead>
<tr>
<th>ALGORITHMS</th>
<th>DECISION TREE</th>
<th>KNN</th>
<th>NAÏVE BAYES</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX</td>
<td>99.9%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>MIN</td>
<td>63.63%</td>
<td>68.32%</td>
<td>49.5%</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>82.28%</td>
<td>81.33%</td>
<td>79.20%</td>
</tr>
</tbody>
</table>

Table-2: Algorithms and their maximum, minimum and average accuracy

**CONCLUSION AND FUTURE WORK**

Prediction of academic performance is very important for any institution from students and for teachers also. It will help us to improve performance of students, we can reduce drop out rate as students with the help of it and we can develop skills among the students and also, we can provide much attention to weak students. So most commonly used algorithm to predict academic performance of students are decision tree, KNN, and Naïve Bayes. And it should be implemented by the institution to predict the performance of students.

![Figure 1: Bar graph of algorithms and their accuracy to predict students academic performance.](image-url)

**REFERENCES**


Efficient Data Mining Techniques to Improve Academic Performance of Students


[18] Mashael Al Iuhaybi, Allan Tucker and Leila Yousefi, "THE PREDICTION OF STUDENT FAILURE USING CLASSIFICATION METHODS: A CASE STUDY"


[21] Najmus Saher Shah, "PREDICTING FACTORS THAT AFFECT STUDENTS’ ACADEMIC PERFORMANCE BY USING DATA MINING TECHNIQUES"


