Efficient Data Mining Techniques to Improve Academic Performance of Students

Rashmi V. Varade[*1], Blessy Thankanchan[2]

[1,2] Jaipur National University, Jaipur

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.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, (2021); DOI :

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 patternevaluation. 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, whichcan discover useful information by analysing data from manyangles or dimensions, categorize that information, and summarizethe relationships identified in the database.[1]

So main objectives of this research are

1) Identifying students whose performance is weak to improve their performance

2) Reviewing data mining techniques such as k-means clustering and decision tree techniques. Naive bias on collected data to get information about students' academic performance.

3) To compare and analysis the results from these techniques and suggest necessary steps to improve students' performance.

4) To find out different reasons that cause students' performance low.

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, mathsgrade, 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-1: shows data mining techniques applied for student's academic performance

Authors	Decision tree	KNN	Naïve Bayes
Nguyen Thai Nghe, Paul Janecek, and Peter Haddawy	84.18%	--	78.57%
Raza Hasan, Sellappan Palaniappan, Abdul Rafi ez Abdul Raziff ,	63.63%	--	90.90%
C. Anuradha1* and T. Velmurugan	--	68.32 %	72.51 %
P.V.Praveen Sundar	--	--	56.3%
V.Ramesh, P.Parkavi, K. Ramar	--	--	49.5%
Yasmeen Shaher Alsalman,Nancy Khamees Abu Halemah, Eman Saleh AlNagi	66 %	--	--
Mashael Al luhaybi, Allan Tucker and Leila Yousefi	76.70%	--	73%
Mashael A.et. al.	--	--	91%
Dr. S. SENTHIL,WONG MU LIN	--	--	88.44%
Mukesh Kumar, Yass Khudheir Salal	98.86%	83.00%	98.00%
Najmus Saher Shah	84%	--	--
R Asif et. al.	83.6	74%	73%
M Koutina et. al.	68.5%	100%	100%
N Goga et. al.	99.9%	--	--
S V. Shinde et. al	97.5%	--	--

DISCUSSION

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

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

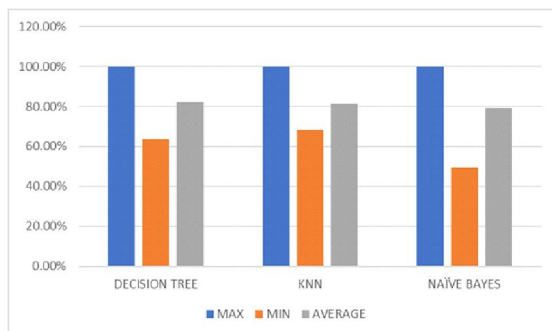| ALGORITHMS	DECISION TREE	KNN	NAÏVE BAYES
MAX	99.9%	100%	100%
MIN	63.63%	68.32%	49.5%
AVERAGE	82.28%	81.33%	79.20%

Figure 1: Bar graph of algorithms and their accuracy to predict students academic performance.

CONCLUSION AND FUTURE WORK

Prediction of academic performance of student is very important for any institution from students and for teachers also. It will help us to improve performance of student, 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, Naïve Bayes. And it should be implemented by the institution to predict the performance of students.

REFERENCES

[1] Abdulmohsen Algarni "Data Mining in Education " IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 6, 2016

[2] Omar Augusto Echegaray-Calderon and Dennis Barrios-Aranibar "Optimal selection of factors using GeneticAlgorithms and Neural Networks for the predictionof students' academic performance "

[3] Brijesh Kumar Baradwaj and Saurabh Pal "Mining Educational Data to Analyse Students Performance"(IJACSA) International Journal of Advanced Computer Science and Applications. 2, No. 6, 2011

[4] Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, The Prediction of Students' Academic Performance Using Classification Data Mining Techniques, Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415 - 6426HIKARI Ltd, www.m-hikari.com http://dx.doi.org/10.12988/ams.2015.53289.

[5] Mashael A. Al-Barrak and Mona S. Al-Razgan, predicting students' performance through classification: a case study, Journal of Theoretical and Applied Information Technology 20th May 2015. Vol.75. No.2.

[6] Edin Osmanbegoviæ and Mirza Suljic, DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[7] Mohammed M. Abu Tair, Alaa M. El-Halees, Mining Educational Data to Improve Students'Performance: A Case Study, International Journal of Information and Communication Technology Research, ISSN 2223-4985, Volume 2 No. 2, February 2012.

[8] Azwa Abdul Aziz, Nor Hafieza Ismailand Fadhilah Ahmad, First Semester Computer Science Students' Academic Performances Analysis by Using Data Mining Classification Algorithms, Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA. (e-ISBN978-967-11768-8-7).

[9] Dipesh Walte , Hari Reddy , Vivek Ugale , Amol Unwane "Overview of algorithms in Educational DataMining for Higher Education: An ApplicationPerspective" International Journal of Engineering Research & Technology (IJERT)Vol. 3 Issue 2, February - 2014 IJERTIJERTISSN: 2278-0181

[10] 'A.Dinesh Kumar, R.Pandi Selvam, K.Sathesh Kumar "Review on Prediction Algorithms in EducationalData Mining " International Journal of Pure and Applied Mathematics Volume 118 No. 8 2018, 531-537ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)

[11] Mamta Sharma and Monali Mavani "Accuracy Comparison of Predictive Algorithms of DataMining: Application in Education Sector"

[12] Nguyen Thai Nghe, Paul Janecek, and Peter Haddawy "A Comparative Analysis of Techniques for Predicting Academic Performance"IEEE Milwaukee, WI37th ASEE/IEEE Frontiers in Education Conference

[13] Raza Hasan, Sellappan Palaniappan, Abdul Rafi ez Abdul Raziff ,Salman Mahmood and Kamal Uddin Sarker.

[14] C. Anuradha* and T. Velmurugan"A Comparative Analysis on the Evaluation of Classification Algorithms in thePrediction of Students Performance"Indian Journal of Science and Technology, Vol 8(15), IPL057, July 2015

[15] P.V.Praveen Sundar "A COMPARATIVE STUDY FOR PREDICTING STUDENT'S ACADEMIC PERFORMANCE USING BAYESIAN NETWORK CLASSIFIERS" IOSR Journal of Engineering (IOSRJEN)e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 2 (Feb. 2013), ||V1|| PP 37-42.

[16] V.Ramesh, P.Parkavi, K.Ramar"Predicting Student Performance:A Statistical and Data Mining Approach"International Journal of Computer Applications (0975 – 8887) Volume 63– No.8, February 2013

[17] Yasmeen ShaherAlsalman, Nancy Khamees Abu Halemah, Eman Saleh AlNagi"Using Decision Tree and Artificial Neural Networkto Predict Students' Academic Performance"2019 10th International Conference on Information and Communication Systems (ICICS)

[18] Mashael Al luhaybi, Allan Tucker and Leila Yousefi"THE PREDICTION OF STUDENT FAILUREUSING CLASSIFICATION METHODS: ACASESTUDY"

[19] Mashael A. Al-Barrak and Mona S. Al-Razgan, predicting students' performance through classification: a case study, Journal of Theoretical and Applied Information Technology 20th May 2015. Vol.75. No.2.

[20] Mukesh Kumar, Yass KhudheirSalal" Systematic Review of Predicting Student's Performance in Academics"International Journal of Engineering and Advanced Technology (IJEAT)ISSN: 2249 – 8958, Volume-8 Issue-3, February 2019

[21] NajmusSaher Shah"PREDICTING FACTORS THATAFFECT STUDENTS' ACADEMIC PERFORMANCEBY USING DATA MININGTECHNIQUES

[22] Raheela Asif, Agathe Merceron, Mahmood K. Pathan, Predicting Student Academic Performance at Degree Level: A Case Study, I.J. Intelligent Systems and Applications, 2015, 01, 49-61 Published Online December 2014 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijisa.2015.01.05

[23] Maria Koutina and Katia Lida Kermanidis, Predicting Postgraduate Students' Performance Using Machine Learning Techniques, L. Iliadis et al. (Eds.): EANN/AIAI 2011, Part II, IFIP AICT 364, pp. 159–168, 2011. © IFIP International Federation for Information Processing 2011

[24] Maria Goga, Shade Kuyoro, Nicolae Goga, A recommender for improving the student academic performance, Social and Behavioural Sciences 180 (2015) 1481 – 1488.

[25] Karishma B. Bhegade and Swati V. Shinde, Student Performance Prediction System with Educational Data Mining, International Journal of Computer Applications (0975 – 8887) Volume 146 – No.5, July 2016.