JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EMPIRICAL STUDY OF PREDICTION MODEL IN EDUCATION DOMAIN

[1]K. P. Raghuvanshi, [2]Dr. V. R. Dhawale, [3]N. P. Pardesi, [4]P. S. Selokar, [5]A. M. Chopade

[1]Professor, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

[2]Professor, K. K. Wagh Institute of Engineering Education and Research, Nashik, India

[3,4,5]Student, Department of MCA, Vidyabharati Mahavidyalaya, Amravati, India

Abstract: Data mining has been used in many different sectors to extract relevant data from the large data sets for analytical purposes. In recent years, several research studies have proven the significance of using data mining techniques in education. With the use of various data mining techniques, relevant data can be efficiently extracted from the large sets of data present in the education system. By analyzing the extracted data, many different aspects regarding education can be predicted with a significantly high accuracy rate. This includes predicting aspects such as newly emerging fields, performance of students and instructors, learning pace of the students, chances of employment after graduation and dropout rate in a particular field etc. This paper presents an empirical study on various papers which show how prediction models can be used to extract the relevant data from the large data sets present in the education system for analyzing several important aspects.

Keywords: **Data Mining, Prediction Model, Data Analysis, Big Data.**

I. INTRODUCTION

Data mining is an exploratory data analysis, an art of extracting information from data, a process of discovering patterns in data, and a task of discovering meaningful data from big data with the aim of obtaining clear and useful results. There are many goals of data mining such as to reduce large data into useful knowledge, to find useful patterns from large data, to identify important trends and previously unknown behavior pattern to find patterns in historical data, and to uncover previously unknown patterns.

II. DATA MINING TECHNIQUES

1. Association

Association analysis is the finding of association rules showing attribute-value conditions that occur frequently together in a given set of data.

2. Classification

Classification is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown.

3. Prediction

Prediction uses a combination of other data mining techniques such as trends, clustering, classification, etc. It analyzes past events or instances in the right sequence to predict a future event.

4. Clustering

Clustering is the task of dividing the data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.

5. Regression

Regression can be defined as a statistical modeling method in which previously obtained data is used to predict continuous quantity for new observations.

6. Sequential Patterns

The sequential pattern is a data mining technique specialized for evaluating sequential data to discover sequential patterns.

III. LITERATURE REVIEW

Sr. No.	Prediction Model	Authors	Techniques	Conference/Journal and Year	Conclusion
1	College Students' Employment	Houwen Fan	Decision Tree	13th International Conference on Intelligent Computation Technology and Automation, 2020	Based on the enrollment, student status management and employment data of a university over the years, this paper preprocesses and analyzes the employment related data, applies the classification algorithm to generate the decision tree model of employment prediction, extracts the prediction rules, and evaluates the accuracy of the employment data Through analyzing and obtaining the regular potential information, the prediction model of student employment and employment unit type is generated. In the process of the experiment, based on the research of predicting the employment of college students, we realized the construction of the decision tree model of employment prediction through programming.
2	Predicting Student Performance	Wan Fairos Wan Yaacob, Syerina Azlin Md Nasir, Wan Faizah Wan Yaacob, Norafefah Mohd Sobri	K-NN Naives Bayes DT – Information Gain DT – Gini Logistic Regression (LR	Indonesian Journal of Electrical Engineering and Computer Science, Dec 2019	In this paper, four supervised data mining algorithms were applied on the students performance data to predict student performance either excellent or non-excellent based on predictive accuracy. The results indicate that the Naïve Bayes classifier outperformed other algorithm compared to Decision Tree, k-NN, and Logistic Regression with accurate and comprehensive classifier.
3	Predict Student's Performance	Annisa Uswatun Khasanah1), Harwati 2)	Bayesian Network & Decision Tree	IOP Conference Series, 2017	Some high influence attributes in student prediction that can be considered by university management to minimized student drop out based on the result of this study are First Semester GPA, First Semester Attendance, Senior High School Department, Gender, Father Occupation, Mother Occupation, Mother Education, and Origin. From this study that feature selection can increase the accuracy rate of the prediction, and Bayesian Network is outperforming Decision Tree since it has higher accuracy rate.
4	Students' Employability Prediction Model	Tripti Mishra., Dharminder Kumar.,	J48 Random forest Random Tree SMO	International journal of applied engineering research, Nov 2016	The aim of this paper was to apply various classifiers to find the employability of students and develop employability model based on the suitable classifier. It was found that J48 algorithm which is

	through Data Mining	Sangeeta Gupt	Multilayer Perceptron Naive Bayes		implementation of pruned C4. 5 Decision Tree algorithm of WEKA is most suitable for the employability prediction.
5	Predict Slow Learners	Parneet Kaura ,Manpreet Singhb ,Gurpreet Singh Josanc	Multilayer Perception 75%, Naïve Bayes 65.13%, SMO 68.42%, J48 69.73% , REPTree	3rd International Conference on Recent Trends in Computing, 2015	Among all data mining classifiers Multi Layer Perception performs best with 75% accuracy and therefore MLP proves to be potentially effective and efficient classifier algorithm. Also comparison of all 5 classifiers with the help of WEKA experimenter is also done, in this case also MLP proves to be best with F-measure of 82%. Therefore, performance of MLP is relatively higher than other classifiers. A model performance chart is also plotted. This research help the institutions to identify students who are slow learners which further provide base for deciding special aid to them.
6	Students' Academic Performance	Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz	Navie Bayes Decision Tree, RB	Applied Mathematical Sciences, Vol. 9, 2015	The amount of data stored in an educational database at IHL is increasing rapidly by the times. In order to get the knowledge about student from such large data and to discover the parameter that contributed to the students' success, the classification techniques are applied to the students' data. This study also conducts a comparative analysis of three classification techniques; DT, NB, and RB using WEKA tool. The experimental result shows that the RB has the best classification accuracy compared to NB and DT.
7	A Study Model On The Impact Of Various Indicators In The Performance Of Students In Higher Education	Jai Ruby1 , K. David2	Chi-square, Information Gain, Gain Ratio, Linear Regression, Correlation.	International Journal of Research in Engineering and Technology	This study paper on performance analysis of student data help the institution to decide on the factors to concentrate for the better performance of the academic results of the students.
8	Performance of Students in Higher Education	Jai Ruby 1 , Dr. K. David 2	ID3, J48, NB Tree, MLP, Simple Cart, Rep Tree, Decision Table	International Journal for Research in Applied Science and Engineering, 2014	This study paper helps the institution to know the academic status of the students in advance and can concentrate on weak students to improve their academic results. All the classification algorithms MLP, ID3, J48, REPTree, NBTree, SimpleCart and Decision Table considered in the study shows prediction accuracy above 68% for the student dataset.
9	Mining Students' Data for Performance Prediction	Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta Gupta	Random Tree, J48	Fourth International Conference on Advanced Computing & Communication Technologies, 2014	This paper focused on identifying attributes that influenced students 'third semester performance. Effect of emotional quotient parameters on placement has been established. Random tree gave higher accuracy of prediction than J 48.
10	Prediction of Student Performance	M. Durairaj #1, C. Vijitha *2	Naïve Bayes Decision Tree	International Journal of Computer Science and Information Technologies, 2014	Using K-Means clustering algorithm, we predicted the pass percentage and fail percentage of the Overall students appeared for a particular examination. The results show the students' performance and it is seems to be accurate. The comparison between Naviebayes algorithm and

					decision stump tree technique shows that the Navie bayes techniques produce accurate result than the other and it is measured using confusion matrix. The results are predicted within 0 seconds.
11	Student Performance	Dorina Kabakchieva	OneR Rule Learner, Decision Tree, Neural Network and K-Nearest Neighbour	International Journal of Computer Science and Management Research, 2012	The classification models, generated by applying the selected four data mining algorithms – OneR Rule Learner, Decision Tree, Neural Network and K-Nearest Neighbour, on the available and carefully pre-processed student data, reveal classification accuracy between 67.46% and 73.59%. The highest accuracy is achieved for the Neural Network model (73.59%), followed by the Decision Tree model (72.74%) and the k-NN model (70.49%). The Neural Network model predicts with higher accuracy the "Strong" class, while the other three models perform better for the "Weak" class.
12	Academic Performance for First Year Student	Ernesto Pathros Ibarra García[1] Pablo Medina Mora[2]	Naïve Bayes	10th Mexican International Conference on Artificial Intelligence, 2011	Predictions for the low and high groups have significant percentage accuracy in some cases, exceeding 70% if the naïve Bayes classifier is used. This shows that it is possible to obtain a good prediction model. It can be used to detect low performing students and high performing students take appropriate decisions even before the courses start and, hence, to revert their academic standing.
13	Early Prediction of Student Success	Zlatko J. Kovačić	CHAID Tree Cart Tree	Informing Science & IT Education Conference, 2010	This study examines the background information from enrolment data that impacts up on the study outcome of Information Systems students at the Open Polytechnic. Based on results from feature selection (Figure 2 and Table 3), the CHAID tree presented in Table 4 and Figures 3 and 4, the CART trees presented in Table 5 and Figures 5 and 6 it was found that the most important factors that help separate successful from unsuccessful students are ethnicity, course programme and course block. Demographic data such as gender and age though significantly related to the study outcome, according to the feature selection result, were not used in the classification trees. Unfortunately the classification accuracy from the classification trees was not very high. In the case of the CHAID tree the overall classification accuracy was 59.4% and in the case of the CART tree slightly higher at 60.5%.

V. CONCLUSION

Prediction models are very important in any domain. We use the prediction model in the Education domain. Using this model we can predict the College Students' Employment, Early Prediction of Student Success, Academic Performance for First Year Student, Performance of Students in Higher Education, Students' Academic Performance and Students' Employability.

IV. FUTURE SCOPE

Data mining techniques are already used for the prediction of success rate of students and chances of employment after graduation. By extracting the relevant data from the database of a particular university or a college, the average number of Start-ups in upcoming years from the respective university or college can be predicted. These techniques can also be used for the prediction of average placement rate of a university or a college in upcoming years.

REFERENCE AND BOOKS

[1] "Data Mining: The Textbook", Charu C. Aggarwal

[2] "Data Mining: Concepts and Techniques", Jiawei Han

[3] "Introduction to Data Mining", Michael Steinbach, Pang-Ning Tan, and Vipin Kumar

[4] "Data Mining Techniques", Arun K. Pujari

[5] Kovačić, Z. (2010), "Early Prediction of Student Success: Mining Students Enrolment Data", Paper presented at Proceedings of Informing Science & IT Education Conference (InSITE) ,Casinio Italia, June, 19-24,2010.

[6] Dorina Kabakchieva, "Student Performance Prediction by Using Data Mining Classification Algorithms", International Journal of Computer Science and Management Research Vol 1 Issue 4 November 2012.

[7] Ramaswami, M., Bhaskaran, R. (2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No.1, January 2010, pp.10-18.

[8] Cortez, P., Silva, A. (2008), "Using Data Mining to Predict Secondary School Student Performance", EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.

[9] Vandamme, J., Meskens, N., Superby, J. (2007), "Predicting Academic Performance by Data Mining Methods", Education Economics, 15(4), pp405-419.

[10] M. Durairaj, C. Vijitha (2014), "Clustering Algorithm for Educational Data Mining: A Systematic Review of Literature and Techniques", 53 International Journal of Scientific Research in Computer Science Applications and Management Studies, Volume 3, Issue 4 (July 2014).

[11] Md. Hedayetul Islam Shovon, "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering,Vol. 2(7), July 2012.

[12] Sajadin Sembiring,"Prediction Of Student Academic Performance by an Application of Data Mining Techniques", International Conference on Management and Artificial Intelligence IPEDR,IACSIT Press, Vol.6,2011.

[13] Tripti Mishra, Dr. Dharminder Kumar , Dr. Sangeeta Gupta , "Mining Students' Data for Performance Prediction", 2014 Fourth International Conference on Advanced Computing & Communication Technologies, 2014 IEEE.

[14] R. R. Kabra, R.R, Bichkar ," Performance Prediction of Engineering Students using Decision Trees", International Journal of Computer Applications, Volume 36, No.11, 2011.

[15] Jai Ruby, K. David, "A Study Model On The Impact Of Various Indicators In The Performance Of Students In Higher Education", International Journal of Research in Engineering and Technology, Volume: 03 Issue: 05 May-2014.