

Performance Based Grouping of Neighbors Students in Progressive Education Datasets

Vijay Anand Sullare
Research Scholar, M.G.C.G.V. Chitrakoot, Satna M.P.

R.S. Thakur
M.A.N.I.T. Bhopal

Bharat Mishra
M.G.C.G.V. Chitrakoot, Satna M.P.

Abstract –
Now a day's the educational organizations are facing the biggest challenges, of the massive growth of educational data. Further they do not have a good policy and to use this data for improving the quality of their managerial decisions in today's scenario. The main goal of higher education institutions is to provide quality of education for their students.
In general the educational database contains the important information for predicting a student's performance, and this Prediction of student's performance in educational environments is of utmost importance. The knowledge mining techniques has provided a decision making tool which can facilitate better resource utilization in terms of students performance. The knowledge mining techniques are more helpful in classifying educational database. In this paper the clustering task is used to assess student's performance from education databases. By using this task we extract the knowledge that can describes students' performance in end semester examination.
Keywords – Educational datasets, knowledge mining, Decision Making, Data Classification, Performance Prediction.

I. INTRODUCTION ABOUT HIGHER EDUCATION

Studies prove that, as compared to school graduates, college graduates have longer life spans, better access to health care, greater economic stability and security, more prestigious employment and greater job satisfaction, less dependency on government assistance, greater knowledge of government, greater community service and leadership, more self-confidence, and less criminal activity and incarceration. Higher education [**1**] improves an individual's quality of living life.
Mining patterns in educational environment is called Educational Data mining, [**2**] concern with designing new techniques to extract new patterns from educational database in order to analyze student's learning behaviours. Lack of complete knowledge in educational system may prevents a better management to achieve quality objectives, knowledge approaches can help bridging this knowledge gaps in higher education system. [12], [13], [14] & [15].

II. RELATED PREVIOUS WORK

Renza Campagni, et. al. (2015) [**3**] presents a data mining methodology to evaluate the careers of University graduated students. It presents different approaches based on clustering and sequential patterns techniques in order to identify strategies for improving the performance of students and the scheduling of exams.

Sajadin Sembiring, et. al. (2011) [**4**] made the presentation on Prediction of student academic performance by an application of data mining techniques. The results of this study reported a model of student academic performance predictors by employing psychometric factors as variables predictors.

J.K. Jothi Kalpana, et. al. (2014) [**5**] made the presentation on 'Intellectual Performance Analysis of Students by Using Data Mining Techniques'. This presentation intends to analysis the student's performance in different categories of measurements.

'A case study in the educational data mining has been given in this paper. It showed how useful data mining can be in higher education particularly to improve graduate students' performance. Finally the Distance-based Approach and Density-Based Approach are used. Each one of these tasks can be used to improve the performance of graduate student.

A paper on 'Application of Data Mining Technique for Prediction of Academic Performance of Student' was given by Bhushan S. Olokar et. al. (2014) [**6**]. This study expressed the strong correlation between mental

condition of student and their final academic performance. Also they clustered the student into group using kernel k-means clustering.

Bhise R.B., et. al. (2013) [**7**] carried out a research on 'Importance of Data Mining in Higher Education System'. The aim of this research was to introduce educational data mining by describing step by step process using technique of K-means (Clustering Methods).

Md. Hedayetul Islam Shovon, et. al. (2012) [**8**] give a presentation on 'Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm'. In this presentation data clustering technique named k-means clustering is applied to analyze student's learning behavior.

Kavita Nagar et. al. (2015) **[9]** presents a paper "Data Mining Clustering Methods:" in this paper gives Clustering is a unsupervised learning method which makes the cluster of objects or documents according to their similarity and dissimilarity bases. This paper gives review about various clustering methods.

M. Durairaj et. al. (2014) **[10]** presents a paper Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. In the paper use k means clustering methods to find pass percentage and fail percentage of the Overall students appeared for a particular examination shows that the Navie bayes techniques produce accurate result than the other and it is measured using confusion matrix.

Oyelade, O. J et. All (2010) **[11]** presents a paper" Application of k-Means Clustering algorithm for prediction of Students' Academic Performance" In this paper, it also implemented k-mean clustering algorithm for analyzing students' result data. It provided a simple and qualitative methodology to compare the predictive power of clustering algorithm and the Euclidean distance as a measure of similarity distance.

III. PROPOSED MODEL FOR CLUSTERING OF EDUCATIONAL DATA.

3.1 Proposed frame work is shown in Figure 1

3.2 : Algorithms for proposed model.

Step 1. Collection of student result data from various engineering institutions.
Step 2. Apply pre-processing techniques to remove noise from data.
Step 3. Divide the data into two parts one is training data (80%) and testing data (20%).
Step 4. Choose appropriate data mining approach to design a model.
Step 5. Trained the model as per given rule base.
Step 6. If the model is not trained, go to step 5.
Step 7. Test the model using Test data.
Step 8. If model is accepted, use it for clustering & prediction purpose.

I. IMPLEMENTATION AND RESULT.

The experimental setup are performed using BE first year Computer Science, Civil Engineering branches student result database collect from engineering institution located in Bhopal city of MP, India and data stored in MS-excel format. This dataset contains continues evaluation marks (i.e. practical, term work, practical quiz, mid-term, test quiz) final marks, grade and locality of student.

Each category of marks normalized in three sub categories A(Average), BP (Blow Average) and AP (Above Average). Clusters are generated using Naïve Bayes Method.

Experiments are done on WEKA 3.7.11. Different branch (Computer Science & Civil Engineering) database are used with different sizes.

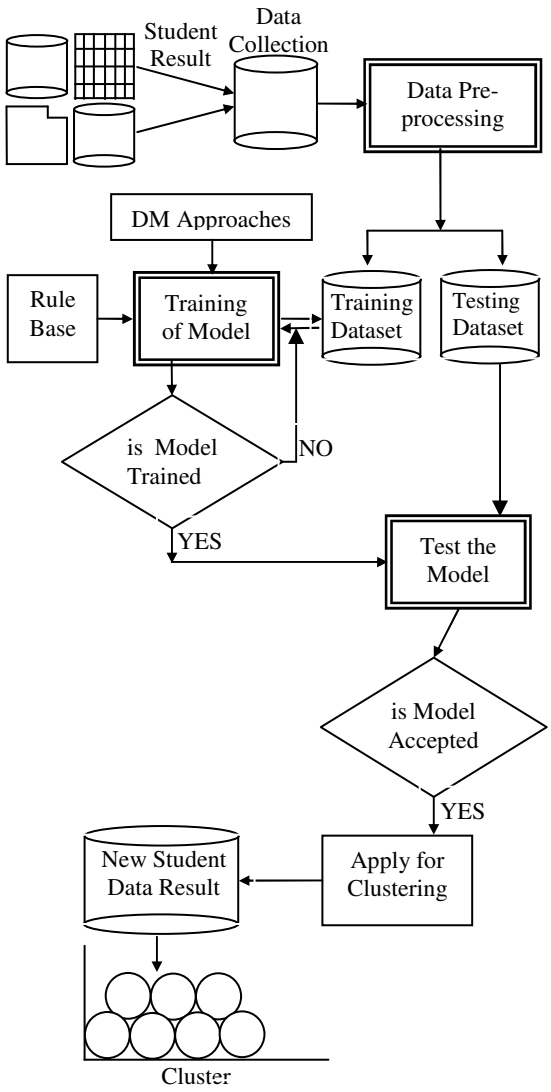

Fig. 1

The experiments are done with 50, 100, 150 and 200 records to Calculate Accuracy Precision, Recall which is shown by Fig. 2, fig. 3 & Fig. 4 respectively.

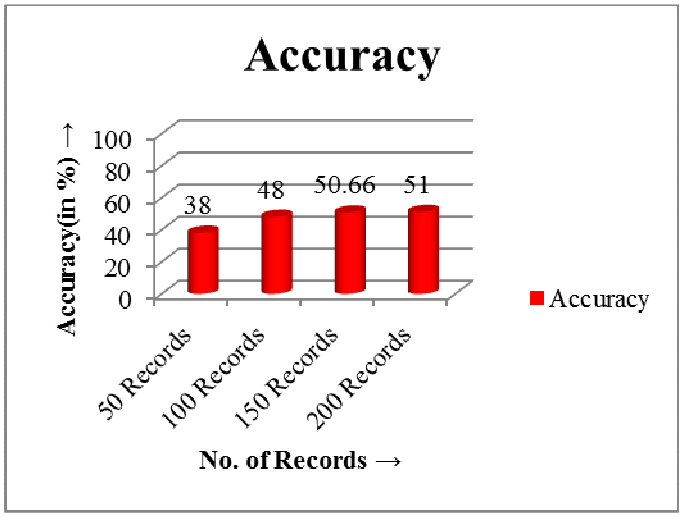

Fig. 2 Accuracy obtained for CS branch records by Naïve Bayes method

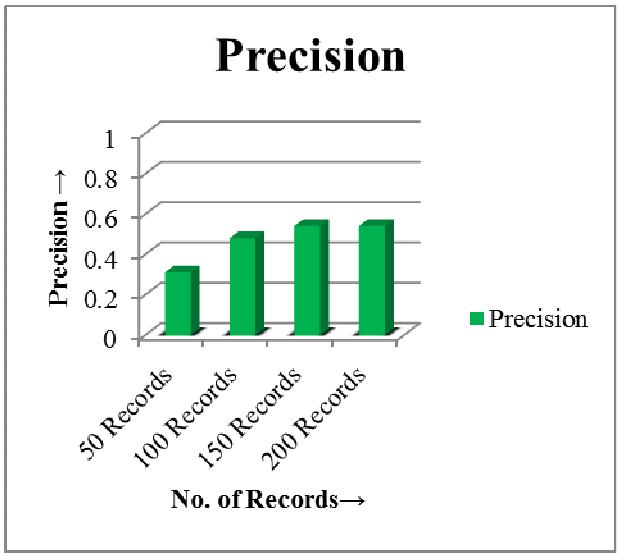

Fig. 3 Precision obtained for CS branch records by Naïve Bayes method

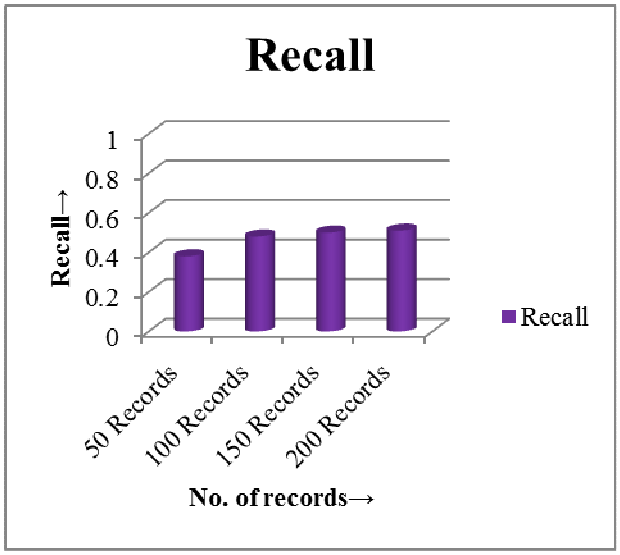

Fig. 4 Recall obtained for CS branch records by Naïve Bayes method

Similarly Accuracy, Precision and recall values are calculate for Civil engineering Branch with 50, 100, 150 & 200 records which is show by fig. 5, Fig. 6 & Fig. 7 respectively.

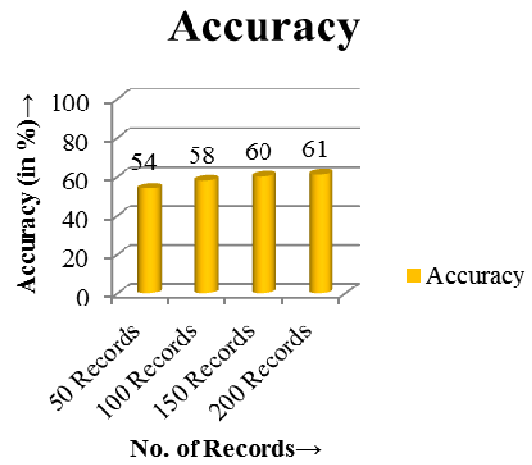

Fig. 5 Accuracy obtained for CE branch records by Naïve Bayes method

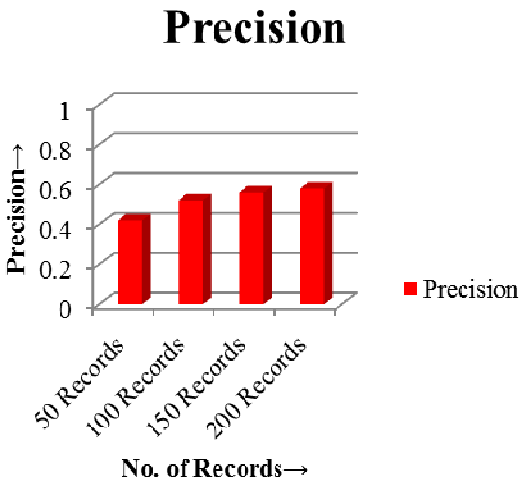

Fig. 6 Precision obtained for CE branch records by Naïve Bayes method

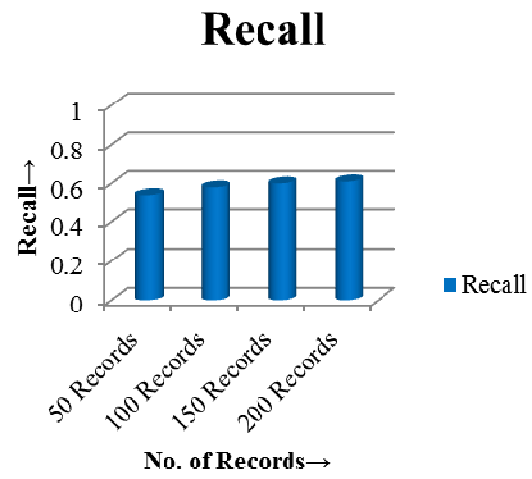

Fig. 7 Recall obtained for CE branch records by Naïve Bayes method

II. CONCLUTION :

This work proposed a framework to classify a knowledge in educational dataset. For this continuous evaluation marks and final grade of B.Tech. first year students of Engineering. Institutions have been considered. The Naïve Bayes algorithm is used for finding centroid of data, which is used for grouping of data and generate the perfect cluster.

The presented work insists the better utilizing of student data in term of getting more insight knowledge which will be help full to find weak point of the student. This framework is used to identify and group the students through their academic performance their used clustering approach to accomplish and since it is best suitable in student dataset.

The discussed approaches also help to identify the performance various among different Branches which gives a open eye to institute management to use their resources accordingly. The used method is scalable in term of the scope i.e. state govt., universities and other regularity bodies may use to get inside the performance of their university affiliating institute. So that they utilize their budget, expertise to improve their colleges.

REFERENCES

[1] http://www.quintcareers.com/college-education-value .
[2] http://www.crosswalk.com/family/homeschool/ why-is-higher-education-important.html.
[3] Renza Campagni, Donatella Merlini, Renzo Sprugnoli, Maria Cecilia Verri "Data mining models for student careers", at Science Direct Expert Systems with Applications, pp 5508–5521,2015,www.elsevier.com.
[4] Sajadin Sembiring, M. Zarlis, Dedy Hartama, Ramliana S, Elvi Wani "Prediction of Student Academic Performance by an Application of Data Mining Techniques" 2011 International Conference on Management and Artificial Intelligence IPEDR vol.6,pp 110-114,2011.
[5] J. K. Jothi Kalpana, K. Venkatalakshmi " Intellectual Performance Analysis of Students by Using Data Mining Techniques" ,International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014, & 2014 IEEE International Conference on Innovations in Engineering and Technology (ICIET'14) On 21st&22ndMarch, Organized by K.L.N. College of Engineering, Madurai, Tamil Nadu, India ISSN (Online) : 2319 - 8753 ISSN (Print) : 2347 – 6710.pp 1922-1929, 2014.
[6] Mr. Bhushan S. Olokar , Prof. Ms. V.M.Deshmukh ,"Application of Data Mining Technique for Prediction of Academic Performance of Student A Literature survey" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 12. PP 3962 – 3965,DEC 2014.
[7] Bhise R.B., Thorat S.S., Supekar A.K., "Importance of Data Mining in Higher Education System", IOSR Journal Of Humanities And Social Science (IOSR-JHSS), ISSN: 2279-0837, ISBN: 2279-0845. Volume 6, Issue 6, PP 18-21 , (Jan. - Feb. 2013), www.Iosrjournals.Org.
[8] Md. Hedayetul Islam Shovon, Mahfuza Haque "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm",International Journal of Advanced Research in Computer Science and Software Engineering , Volume 2, Issue 7, ISSN: 2277 128X, pp 353-355, July 2012,
[9] Kavita Nagar "Data Mining Clustering Methods: A Review" International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Volume 5, Issue 4, ISS: 2277 128X, pp 575-579, 2015, www.ijarcsse.com.
[10] M. Durairaj, C. Vijitha. "Educational Data Mining for Prediction of Student Performance Using Clustering Algorithms" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, pp 5987-5991.
[11] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C "Application of k-Means Clustering Algorithm for Prediction of Student's Academic Performance", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 1, 2010, pp 292-295.
[12] Brijesh Kumar Baradwaj, Saurabh Pal "Mining Educational Data to Analyze Student's Performance" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, pp 63-69,2011.
[13] http://scialert.net.
[14] J. Han and M. Kamber, "Data Mining: Concepts and techniques", Morgan Kaufmann Publishers, Elsevier India, 2001.
[15] A. K. Pujari, Data Mininng Techniques, University Press 2001.