



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

# Data Mining Technique for Prediction of Academic Performance of Student using SOM

Bhushan S. Olokar<sup>1</sup>, Dr. V. M. Deshmukh<sup>2</sup>

ME<sup>4th</sup> Sem, Dept. of Information Technology, Prof. Ram Meghe Institute of Technology & Research,  
Badnera-Amravati, India<sup>1</sup>

Associate Professor & HOD, Dept. of Information Technology, Prof. Ram Meghe Institute of Technology & Research,  
Badnera-Amravati, India<sup>2</sup>

**ABSTRACT:** Datamining is the technique that can be directed to support the specific need of each of the participants in the education system and the process. Right now there are so many prediction systems are available with difference approach and techniques in student performance prediction that has been reported by researcher. The knowledge is hidden among the educational datasets that is extractable through data mining techniques. This paper designed to clarify the capability of data mining techniques in the context of lower education to higher education in university or other educational system by taking in mind different approaches. In this research, the classification task is use to evolved student performance and there are so many approaches to figure out the performance of student.

By this different method we can extract knowledge that describes student performance, It definitely helps to improve the performance of student.

**KEYWORDS:** Data mining, SOM, kernel-k-means, Student Performance, feature Vector.

### I. INTRODUCTION

The advantage of Information technology in various fields has lead the large volume of data storage in various format like record, field, documents and many new data formats the data collection from different application requires proper method of extracting knowledge from large repositories for better decision making. KDD often called data mining aims at the discovery of useful information from large collection of data [1]. The main function of data mining is applying various methods and algorithms in order to discover and extract pattern of stored data [2]. Data mining and the knowledge discovery application have got a reach focus due to its significant in decision making and it has become and essential component in various organisations. Data mining technique have been introduced into new field of database, Machine learning, Pattern Reorganization, Artificial Intelligence and Computational capabilities etc. These are increasing research interests in data mining education, this new emerging field knows as educational data mining

There are increasing research interests in education field using data mining. Data mining techniques concerns to develop the methods that discover knowledge from data and used to uncover hidden or unknown information thatis not capable of being seen, but stronglyuseful [3]. The data can be personal or academic which can be used to understand student'sbehaviour to improve coursework, to improve teaching and many other benefits.

The topic of prediction of academic performance is researched widely. The prediction of student success in order institution is still the most topical debates in higher study centres. In the previous studies, the model of Tinto [3] is the predominant theoretical framework for considering consisting factors in academic goal. The model of Tinto's considers the process of student attrition as a psychological interplay between the characteristics of the student entering university and the experience at the institute. The use of data mining techniques in this field is relatively advanced. There are many data mining techniques was used in this field, such as decision tree, neural networks, Bayesian network, and so on [4].

This study investigates the educational domain of data mining using a case study from data that come from student's behaviour. It showed what kind of data could be collected, how could we pre-process the data, how to apply kernel method in data mining on the data, and finally how can we benefited from the discovered knowledge. In this study, university students were predicated his/her class test grade by using SOM classification and grouped the students



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

according to their similar characteristics, forming clusters. The clustering process was carried out using kernel k-means algorithm technique.

## A. KERNEL K-MEANS FOR CLUSTERING:

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different the objects in another group [4]. In educational area, clustering will be used to grouping students according to their behaviour and performance. In this study we used Kernel K-means algorithm to cluster the given data. A drawback to original K-means is that it cannot separate cluster that are non-linearly separable input space. Kernel K-Means is one approaches have emerged for tackling such a problem. Kernel K-means, where, before clustering, points mapped to a higher dimensional feature space using a non-linear function, and then Kernel K-means partitions the points by linear separator in new space[13]. Kernel K-means has been extended to efficient and effective large scale clustering [4], since the original Kernel K-means had serious problems, such as the high clustering cost due to the repeated calculations of kernel values, or insufficient memory to store the kernel matrix, that make it unsuitable for large corpora. The new clustering scheme is a large scale clustering for Kernel K-means algorithm [4].

## B. SELF-ORGANIZATION MAP:

The research work has been motivated by a number of practical data mining projects Where SOM has been a central data analysis tool [6]. It has become apparent that while the SOM can be used to quickly create a qualitative overview of the data, turning this qualitative information to quantitative characterizations requires a great deal of expertise and manual work. There is no wide consensus or understanding of the methods needed for post-processing of the SOM-based data analysis. The subsequent research has concentrated on devising such methods and on gaining a better understanding of the possibilities, strengths, and weaknesses of the SOM in data exploration.

## II. LITERATURE REVIEW

Bharadwaj and Pal [6] base their experiment only on Previous Semester marks, seminar performance, Assignment, class test marks, attendance, Lab work to predict end semester marks. Records of 50 students of Session 2007 to 2010 MCA of Purvanchal University were considered. The paper calculates Split info, gain ratio of each predictor and products prediction rules.

The students dropping out of an open polytechnic of New Zealand due to failure has been explored by Kovaic.Z [10]. Enrolments data consisting of socio-demographic variables (age, gender, ethnicity, education, work status, and disability) and study environment (course programme and course block), of 435 students of polytechnic students of Information system course were collected. The final label consisting of two categories PASS (those who completed the course) and FAIL (Those who did not complete) were considered. Feature selection indicated that most important attributes for prediction are ethnicity, course programme and course block.

Most cited literature survey in educational Data mining have been by Romero and Ventura [1] which indicate performance prediction as one of the emerging field of educational data mining Various Bayesian Classification techniques have been Used and comparative study suggest that Ensemble methods gives best overall accuracy.

Cheewaparakobkit [14] considered 1600 students records bet 2001 and 2011 in Thailand University and applies decision tree and neural network to most important factors affecting student's academic achievement. Decision tree proves to be a better classifier than the neural network with 1.31% more accuracy. Number of hours worked per semester, additional English course, no of credits enrolled per semester and marital status of the students are major factors affecting the performance.

[9] applied the classification as data mining technique to evaluate student performance; they used decision tree method for classification. This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising. [10] applied the classification as data mining technique to evaluate students' performance; they used decision tree method for classification. This study allows the University management to prepare necessary resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to concentrate upon in higher education systems for support. [11] applied the



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

association rule mining analysis based on students' failed courses to identify students' failure patterns. The goal of their study is to identify hidden relationships between the failed courses and suggest relevant causes of the failure to improve the low capacity student's performances.

[12] Used k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as students. Used Bayesian Classification Method as a data mining technique and concluded that students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students' other habits, family annual income and students' family status were highly correlated with the student's academic performance. Used simple linear regression analysis and it was found that the factors like mother's education and student's family income were highly correlated with the student's academic performance. Conducted a study on the student's performance using association rule technique and they find the interestingness of student in opting class teaching language.

### III. PROPOSED WORK

Educational data mining is an emerging field regarding the prediction of future performance. The objective of the proposed methodology is to build a classification model that classifies a student's performance and has been built by combining the Standard Process for Data Mining that includes: business understanding, data understanding, data preparation, modelling and finally application of data mining techniques which is classification in the present study. Particularly, we will implement the rules into SOM algorithm to predicate the students' final grade. Also, we clustered the student into groups using kernel k-means clustering. This study expressed the strong correlation between the mental condition of the student and their final academic performance. Data Mining Technique has a potential in performance monitoring of universities and other levels of education, offering historical perspectives of students' performances. The results may both complement and supplement tertiary education performance monitoring and assessment implementations.

#### A. CLASSIFICATION ALGORITHMS

The classification and clustering algorithms are used. Classification consists of predicting a certain outcome based on a given input. To predict the outcome, the algorithm processes a training set containing a set of attributes and their respective outcomes. It tries to discover relationships between the attributes that would make it possible to predict the outcome. Further, the algorithm is given a data set not seen before which contains the same set of attributes, except for the prediction attribute – not yet known. It analyses the input and produces a prediction.

The accuracy for prediction defines how "good" the algorithm is. Main algorithms for classification are ID3 and C4.5. ID3 algorithm was originally developed by J. Ross Quinlan at the University of Sydney. ID3 algorithm induces classification models, or decision trees, from data. It is a supervised learning algorithm that is trained by examples for different classes. After being trained, the algorithm should be able to predict the class of a new item. ID3 identifies attributes that differentiate one class from another. All attributes must be known in advance, and must also be either continuous or selected from a set of known values. For instance, temperature and country of citizenship are valid attributes. To determine which attributes are the most important, ID3 uses the statistical property of entropy. It measures the amount of information in an attribute. This is how the decision tree, which will be used in testing future cases, is built. Imagine that you have a dataset with a list of predictors or independent variables and a list of targets or dependent variables. Onward, by applying a decision tree like J48 on that dataset would allow you to predict the target variable of a new dataset record [15]. Decision tree J48 is the implementation of algorithm ID3 (Iterative Dichotomiser 3) developed by the WEKA project team. R includes this nice work into package RWeka [15]. Data Clustering is the process of making a group of abstract objects into classes of similar objects [16]. A cluster of objects of data can be treated as one group.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

## IV. EXPERIMENTAL RESULT

In the proposed work we are going to create a prediction System by using two algorithms k-means & SOM for this we need to generate future vector from the row datasets in the form of clusters.

Following procedure followed for the analysis of proposed Solution:

- The row Datasets of the student is obtained for the processing It undergoes the cleaning Procedure.
- Clustering algorithm is applied on the data, while row data is also maintained.

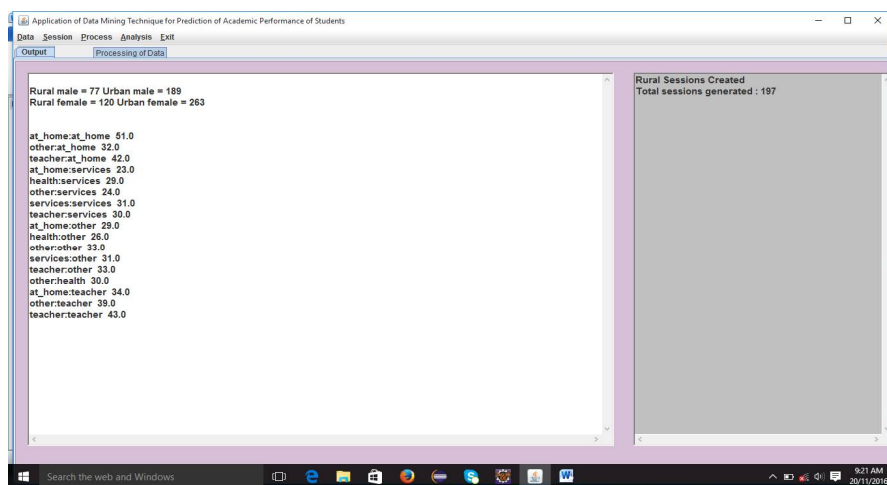


Fig1. Classification of dataset

- Classification algorithm is applied on clusters to generate the iterations by using this classification on sorted data set we can analysis the prediction of student performance.
- K-means Clustering

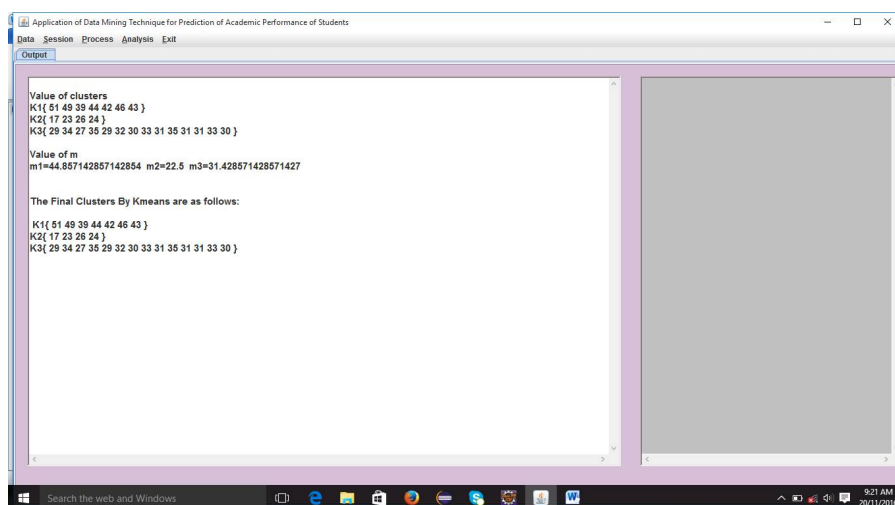


Fig2. Clustering using K-means

- Final Data Analysis

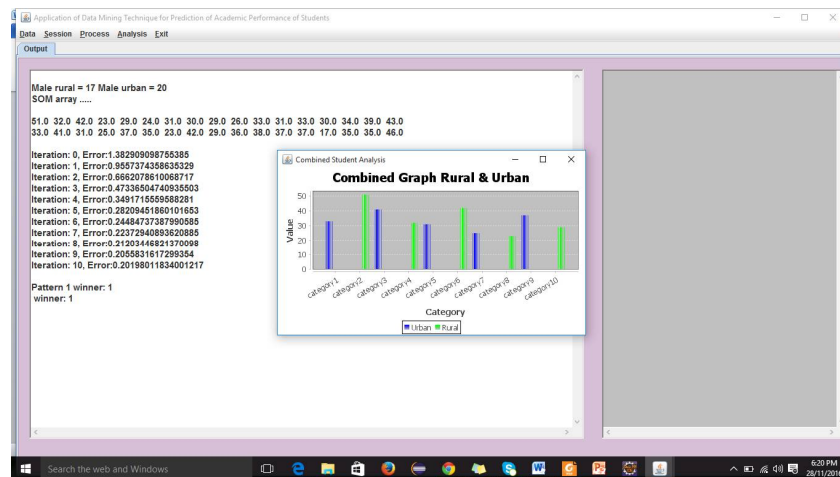


Fig3. Data analysis by using SOM array

The proposed method is based on the parameter Rural and Urban. The parameter observed in both cases of student towards the incremental prediction. The proposed System will definitely helpful to calculate the future prediction of the Student performance.

## V. CONCLUSION

The classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the students previous database, to predict the performance at the end of the semester. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination.

## REFERENCES

- Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.
- U. Fayyad, piatesky, G. Shapiro, and P. Smyth, from data mining to knowledge discovery in database, AAAI Press / The MIT press, Assachuseets Institute of Technology. ISBN 0-262 56097-6,1996.
- V. Tinto, "Limits of theory and practice in student attrition," Journal of Higher Education no. 53, pp. 687-700, 1982.
- Romero, C. and Ventura, S., "Educational Data mining: A survey from 1995 to 2005", Expert systems With Application" (33) 135-146. 2007
- B.K.Bhardwaj and S.Paul , "Mining Educational Datato Analyze Students Performance", InternationalJournal Advanced Computer Science and applicationVol. 2 No. 6 , 2011 .
- Kohonen, T., "Self-Organizing Maps", Series in Information Sciences, second edn. 1997, Springer, Heidelberg.
- Y.J. Lee. And O.L Mangasarian, " A Smooth Support Vector Machine for classification", Journal of Computational Optimization and Applications.20, 2001, pp.5-22
- Kovaic Z (2010) "Early prediction of student success; mining student enrolment data" paper presented at proceeding if information science &IT Education and conference (InSITE), casinio italia, june, 19-24,2010
- Christopher Burges. "A Tutorial on support vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2(2), 1998
- Mark Girolami." Mercer Kernel Based Clustering in Feature Space" I EEE Trans. On Newral Networks.
- P.Cheewaprabokkit, "Study of Factor Analysis AffectingAchievements of Undergraduate", Paper presented at International Multi Conference of Engineers and ComputerScientists, IMECS , Hong Kong, HK, March 13 - 15, 2013.
- Furqan,M.,A.Embong, Suryanti,A, Santi W.P., Sajadin,S., "Smooth Support Vector Machine For Face Recognition Using Principal Componen Analysis". Proceeding 2<sup>nd</sup>International Conference On Green Technology and Engineering (ICGTE), 2009. Faculty of Engineering Malahayati University, Bandar Lampung, Indonesia.



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

## International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 11, November 2016**

13. Dr. Neeraj Bhargava, Girja Sharma, Dr. Ritu Bhargava, Manish Mathuria. "Decision Tree Analysis on J48 Algorithm for Data Mining". International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013.
14. Aastha Joshi, Rajneet Kaur. "A Review: Comparative Study of Various Clustering Techniques in Data Mining". International Journal of Advanced Research in Computer Science and Software Engineering. Volume 3, Issue 3, March 2013.
15. B. Sen, E. Uçar and D. Delen, "Predicting and Analyzing Secondary Education Placement-Test Scores: A Data Mining Approach", Expert System with Application, Volume 39, Issue 10, 2012.
16. T. Nghe, J. Paul, Aneek and Peter Heddawy, "A Comparative Analysis of Techniques for Predicting Academic Performance", Paper presented at 37<sup>th</sup> ASEE/IEEE Conference, Frontiers in Education Conference- Global Engineering: Knowledge Without Borders, Opportunities Without Passports, Milwaukee, WI, October 10-13, 2007.