Student Progress Predictor

R. Ksohy
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai-400058, India

A.V. Sakpal
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai-400058, India

R. More
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai-400058, India

ABSTRACT

Using data mining algorithms can help discover-ing pedagogically relevant knowledge contained in databases obtained from Web-based educational systems. These findings can be used both to help teachers with managing their class, understand their students learning and reflect on their teaching and to support learner reflection and provide proactive feedback to learners.

Keywords
EDM, ANN,WEKA

1. INTRODUCTION
1.1 Data mining
Data mining is a process of taking out useful information and patterns from large amount of data. It is also called knowledge discovery process, knowledge/pattern analysis or knowledge mining from data. The main goal of this analysis process is to take information from a data set and convert it into an understandable and meaningful structure for further use. Data Mining is used for solving problems by analyzing data that is present in the databases.

1.2 Educational Data mining
Educational Data Mining (EDM) is a process which is concerned with developing various techniques or methods for extracting the different types of data that come from educational settings, and use of those methods for better understanding of students. The main area of EDM is analyzing students performance. Another important field is mining enrolment data. Main uses of EDM include student performance prediction and studying students learning to suggest improvements in current educational practice. EDM can be taken as one of the learning sciences and as a field of data mining.
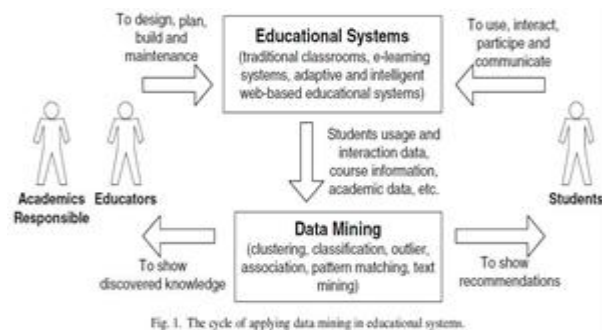

Fig. 1. The cycle of applying data mining in educational systems.

Fig. 1. Shows process of data mining in Educational System.

2. LITERATURE SURVEY
[1]We have studied there is a stronger need to develop algorithms that have various applications of data mining like analysis and higher accuracy and less processing time. To develop new visualization of data, grouping of students, student algorithms there is a lot of scope in this field. A lot of work performance analysis and planning and scheduling etc. and is being done in this area by researchers. these can be implemented using various algorithms of classification and clustering techniques. [2]The concept of data mining and surveyed that at present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. They found that commercial, educational and scientific applications are increasingly dependent on these methodologies. [3]The method known as K-mean clustering, it calculates initial centroids instead of random selection, due to which the umber of iterations is reduced and elapsed time is improved. [4]Kmeans clustering algorithm for the prediction of Students Academic Performance. The ability to monitor the progress of students academic performance is a critical issue to the academic community of higher learning. This paper is aims to present a systematic review on different clustering techniques applied for educational data mining to predict academic performance of students and its implications.

3. ALGORITHM
3.1 K-Means Clustering Algorithm
K-mean clustering algorithm, clusters are fully dependent on the selection of the initial cluster centroids. K data elements are selected as initial centers and then the distances of all data elements are calculated by Euclidean distance formula. Data elements having less distance to centroids are moved to the appropriate cluster. The process is continued until no more changes occur in clusters. K-mean clustering algorithm steps:
INPUT: Number of desired clusters K Data objects D= d1, d2...dn OUTPUT: A set of K clusters Steps:

1) Randomly select k data objects from data, set D as initial centers.
2) Repeat;
3) Calculate the distance between each data object
   di $(1 \le i \le n)$ and all k clusters C j$(1 \le j \le k)$ and assign data object di to the nearest cluster.

4) For each cluster j $(1 \le j \le k)$, recalculate the cluster center.

5) Until no change in the center of clusters.

6) Time complexity of K-mean Clustering is represented

7) by O(nkt)

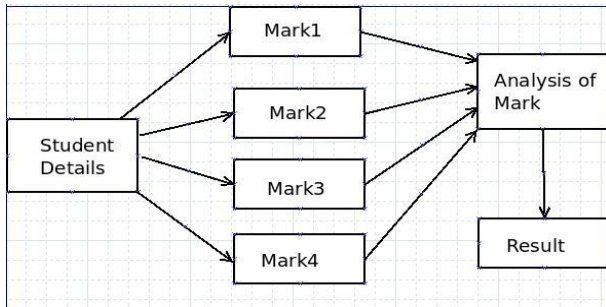Note: Where n is the number of objects, k is the number of clusters and t is the number of iterations.

Fig. 2. It shows Student mark analysis.

4. SCOPE OF PROJECT

1) Predicting students future learning behaviour: here we are classifying the student on the basis of their behaviour also .so the teacher can see and compare their behaviour and can easily understand that student.

2) Discovering or improving domain models: by making use of previously available database we create well structured and categories database of student.

3) Providing Feedback: teacher and student can easily interact with each other and teacher can even give his suggestions or feedback to the student.

4) Detecting undesirable student behaviours: teacher will know all the details of the student as well as in which category he is classified so he will get idea how the student is behaving now and might behave in future.

5) New way of dividing Student: most of the system divide students on the basis of marks only, here we are considering not only marks as well as other criteria like attendance, lab-work, assignment marks, behaviour etc.

5. REQUIREMENTS
5.1 Software Requirements

1) Xampp server to create database.
2) Student Database.
3) OS : Windows. Linux.

5.2 Hardware Requirements

1) Pentium Processor
2) 1 GB RAM

6. IMPLEMENTATION
6.1 Using WEKA Tool

6.1.1 Introduction

WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. WEKA is a state-of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms

for data preprocessing, classification, regression, lustering, association rules; it also includes a visualization tools.

There are four options that can be used :

1) Simple CLI provides a simple command-line interface and allows direct execution of Weka commands.
2) Explorer is an environment for exploring data.
3) Experimenter is an environment for performing experiments and conducting statistical tests between learning schemes.
4) KnowledgeFlow is a Java-Beans-based interface for setting up and running machine learning experiments.

Fig. 3. Shows WEKA Explorer starting Window.
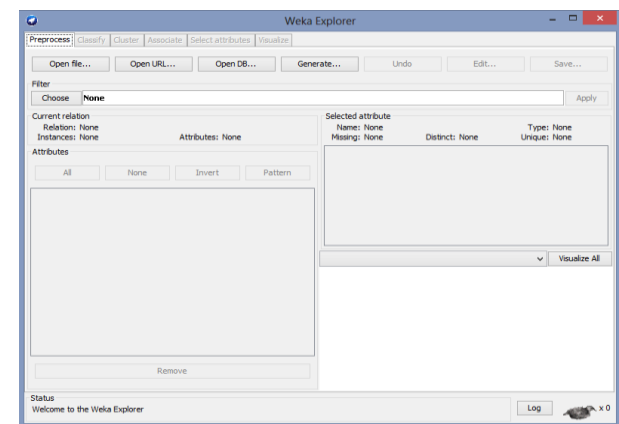
WEKA Explorer Window

Fig. 4. Shows WEKA Exploration process

Load the data into Preprocess window :
Once the data is loaded, WEKA recognizes attributes that are shown in the Attribute window.
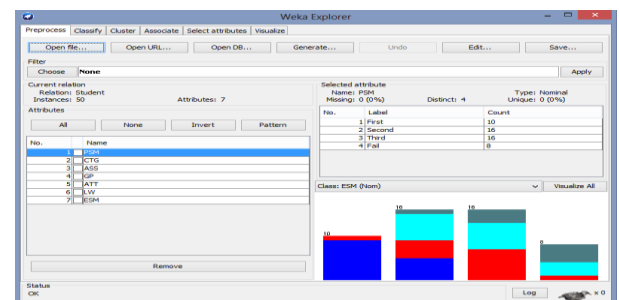

Fig. 5. Shows process Visualization Window.

You can select a class in the Class pull-down box. You can Visualize the attributes based on selected class. One way is to visualize selected attribute based on class selected in the Class pull-down window, or visualize all attributes by clicking on Visualize All button.
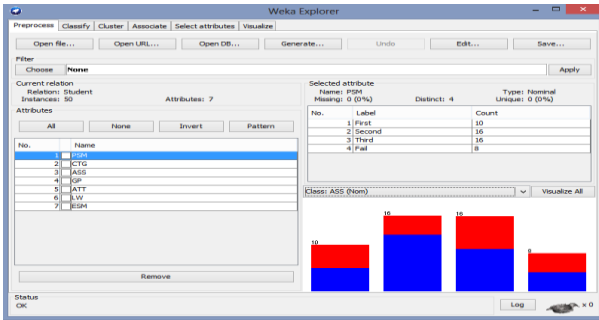

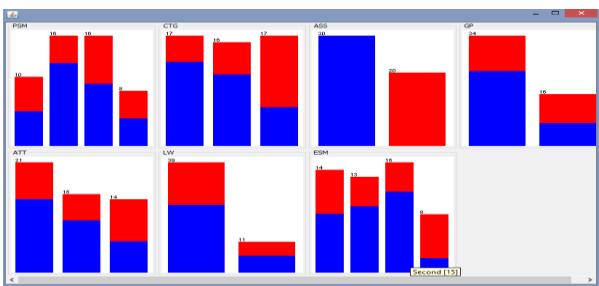**Fig. 6. Shows process Visualization Window**

Fig. 7. Shows process Visualization Window.

Pre-processing tools in WEKA are called filters. WE KA contains filters for discretization, normalization, resampling, attribute selection, transformation and combination of attributes. Apply the Discretization filter and output of Discretization filter is :
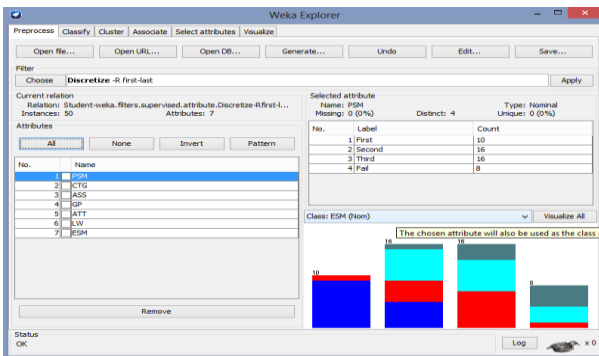
Fig. 8. Shows process Visualization Window.

6.1.2 Classifiers

Classifiers in WEKA are the models for predicting nominal or numeric quantities. The learning schemes available in WEKA include decision trees and lists, instance-based classifiers, support vector machines, multilayer perceptrons, logistic regression, and bayes nets. Meta classifiers include bagging, boosting, stacking, error-correcting output codes, and locally weighted learning.

The Classifiers window look like :
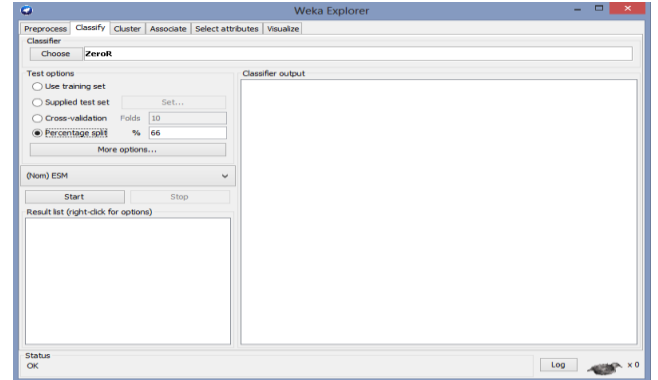
Fig. 9. Choose K-means Clustering Algo.

In classifiers select Tree rule in that select sub-rule is J-48 and then start Classification. The Classification is look like :
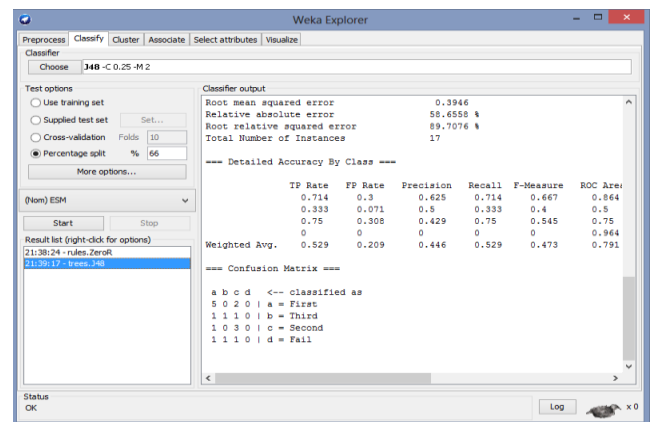
Fig. 10. Shows process of Classification window.

6.1.3 Visualisation of Results

After training a classifier, the result list adds an entry. Right-click on the entry in Result list for which you would like to visualize a tree.
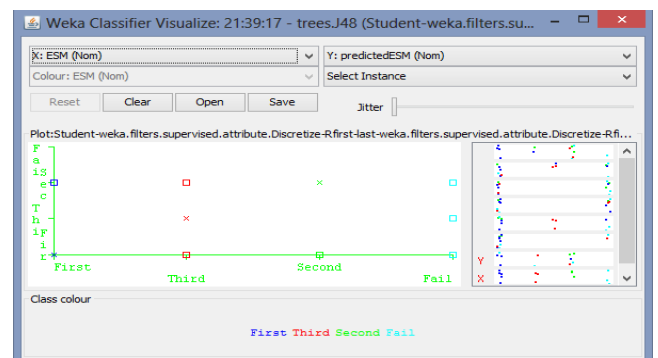

Fig. 11. Shows process of centroids

6.1.4 Clustering Data

WEKA contains clusters for finding groups of similar instances in a dataset. The clustering schemes available in WEKA are k-Means, EM, Cobweb, X means, Farthest First. Clusters can be visualized and compared to true clusters (if given). Evaluation is based on log likelihood if clustering scheme produces a probability distribution.
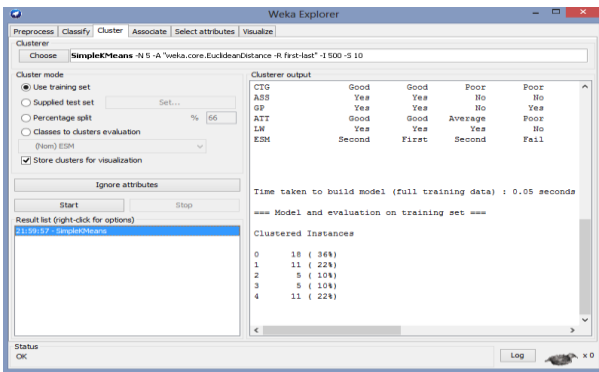Analyze it with k-means clustering scheme.

Fig. 12. Shows process of classifier window.

6.1.5 Visualization of Results(on Clustering Data)
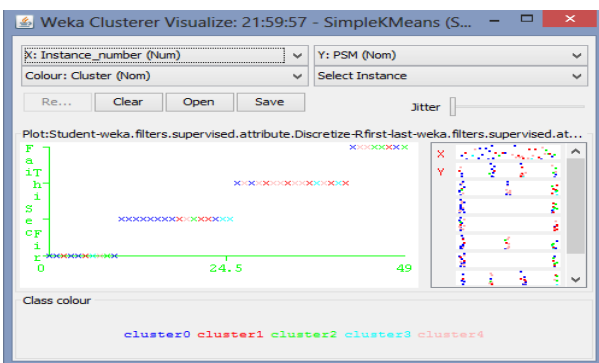Another way of representation of results of clustering is through visualization.

Fig. 13. Shows process of centroid.

6.2 Using Web Programming
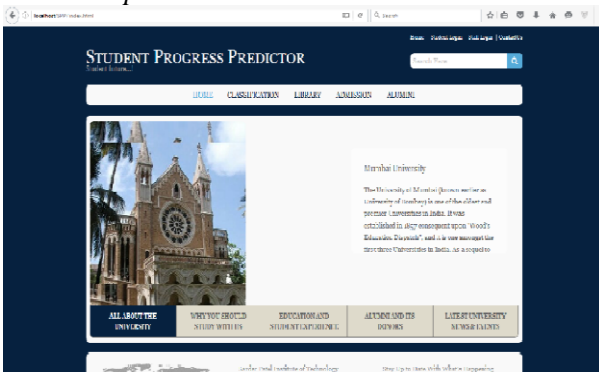
6.2.1 Output

Fig. 14. Shows home page

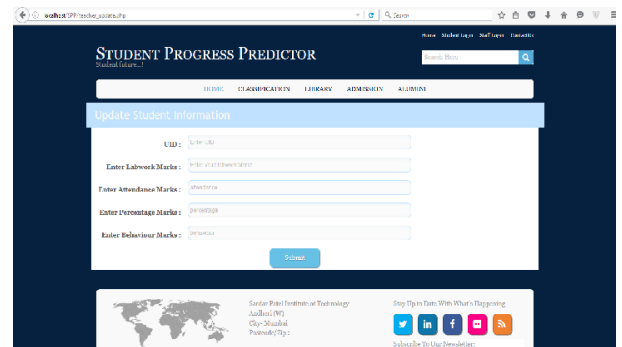Update Marks of Student using four parameters:

Fig. 15. Shows process of marks Updation

Classified Student in Various class :
Class A :

Fig. 16. Shows process of Classified Table(class A)

Class B :

Fig. 17. Shows process of Classified Table(class B)

Class C :

Fig. 18. Shows process of Classified Table(class C)

7. CONCLUSION

Using K-Means clustering algorithm, we will classify the students in 3 different categories on the basis of percentage, attendance, assignment, lab work and behaviour of the students appeared for a particular examination. So the teacher can easily identify the need of the student Because of this student as well as teacher can easily communicate with each other.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Han, J. M. Kamber, Data mining: concepts and techniques, San Francisco: Morgan Kaufman (2012).

[2] Agrawal, R. R. Srikant. ”Fast Algorithms for Mining Association Rules” in Proceedings of VLDB, Santiago, Chile (2010).

[3] SPSS, Clementine, www.spss.com/clementine/ (accessed 2011).

[4] SODAS,http://www.ceremade.dauphine.fr/touati/sodaspa gegarde. html (accessed 2013)

[5] Witten, E. Frank, Data Mining, Practical Machine LearningTools and Techniques with Java Implementation, Morgan Kaufmann Publishers, 2010.

[6] R. Kirkby, WEKA Explorer User Guide for version 3-3-4, University of Weikato, 2012.

[7] M. H. Dunham, Data Mining, Introductory and Advanced Topics, Prentice Hall, 2012.