# **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. 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-Moons Clustering Algor

# 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  $(1 = i_i=n)$  and all k clusters C  $j(1 = j_i=k)$  and assign data object di to the nearest cluster.
  - For each cluster j (1 = ji=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.
- Discovering or improving domain models: by making use of previously available database we create well structured and categories database of student.
- 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 :

- 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

•	Weka	Explorer		- • ×
Preprocess Classify Cluster	r Associate Select attributes Visualize			
Open file O	Ipen URL Open DB Gene	rate Un	io Edit	Save
Filter				
Choose None				Apply
Current relation Relation: None Instances: None	Attributes: None	Selected attribute Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes				
All No	ne Invert Pattern			
				1
				<ul> <li>Visualize All</li> </ul>
	Remove			
Status Welcome to the Weka Explore	er			Log x

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.

🗢 Weka	a Explorer	- 🗆 🗙
Preprocess Classify Cluster Associate Select attributes Visualize		
Open file Open URL Open DB Ger	nerate Undo Edit	Save
Choose None		Apply
Current relation Relation: Student Instances: 50 Attributes: 7	Selected attribute Name: PSM Missing: 0 (0%) Distinct: 4	Type: Nominal Unique: 0 (0%)
Attributes	No. Label	Count
All None Invert Pattern	1 First 2 Second	10 16
No. Name	3 Third	16
>         ki0           4         (i)0           5         (i)1           7         (i)1           9         (i)1           9         (i)1           9         (i)1           9         (i)1           10         (i)1           11         (i)1           12         (i)1           13         (i)1           14         (i)1           15         (i)1           15         (i)1           15         (i)1           16         (i)1           17         (i)1           16         (i)1           17         (i)1           16         (i)1           17         (i)1           18         (i)1           17         (i)1           18         (i)1           18 <td< th=""><th>Cisso: A55 (Verry)</th><th>Visualize Al</th></td<>	Cisso: A55 (Verry)	Visualize Al
Status		
OK		Log

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 :

3		Weka Explorer	
Preprocess Classify	Cluster   Associate   Selec	attributes Visualize	
Classifier			
Choose ZeroR			
Test options		Classifier output	
<ul> <li>Use training set</li> </ul>			
<ul> <li>Supplied test set</li> </ul>	Set		
O Cross-validation	Folds 10		
Percentage split	% 66		
Mon	e options	7	
(Nom) ESM		✓	
Start	Stop		
Result list (right-dick fr	vr ontions)		
Status			

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 :

<i>•</i>			vveka Ex	plorer				
Preprocess Classify	Cluster Associate	Select attributes Visu	alize					
Classifier								
Choose J48 -C	0.25 -M 2							
Test options		Classifier output						
<ul> <li>Use training set</li> </ul>		Root mean squa	red error		0.39	46		
Supplied test set	Set	Relative absol	ute error		58.65	58 %		
0		Root relative	squared e	rror	89.70	76 🐧		
Cross-validation	Folds 10	Total Number o	f Instanc	es	17			
<ul> <li>Percentage split</li> </ul>	% 66	Detailed 3	COURSON B	Class	_			
More opt	tions	Decarred P	couracy p	y crass	-			
			TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Are
Nom) ESM			0.714	0.3	0.625	0.714	0.667	0.86
			0.333	0.071	0.5	0.333	0.4	0.5
Start	Stop		0.75	0.308	0.429	0.75	0.545	0.75
Result list (right-click fr	or options)		0	0	0	0	0	0.964
21:38:24 - rules. ZeroR	1	Weighted Avg.	0.529	0.209	0.446	0.529	0.473	0.791
1:39:17 - trees.348		=== Confusion	Matrix ==					
		abcd <	classifi	ed as				
		5020 a=	First					
		1110   b -	Third					
		103010-	Second					
		1110 a	rall					
		1			_			,
status							1.00	
рк							Log	-428

Fig. 10. Shows process of Classification window.

#### 6.1.3 Visualisation of Results

After training a classifier, the result list adds an entry. Rightclick 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.

•					v	/eka Explo	orer			-		
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize							
Clusterer												_
Choose	Simp	leKMea	ns-N5-A"∨	veka.core.Euclidea	nDistance -	R first-last" -	I 500 -5 10					
Cluster mod	le				Cluster	er output						-
Use tra	ining set				CTG		Good	Good	Poor	Poor		^
Supple	d test set		Set		ASS		Yes	Yes	No	No		
0.0				n/ [cc	GP		Yes	Yes	No	Yes		
Oreiten	tage spar			78 00	TW		Yea	Yea	Xverage	POOL		
Classe	s to cluste	irs evalua	tion		ESM		Second	First	Second	Fail		
(Nom)	ESM			$\sim$								
Store of	lusters fo	r visualiz	ation									
		Ionore a	ttributes		1							
		-9			Time	taken to	build model	(full tra	ining data)	: 0.05 #	conds	
	Start			Stop								
Result list (r	ight-click	for option	18)		;	NOUEL and	evaluation	on trainin	ig sec			- 11
21:59:57 - 5	SimpleKMe	ans			Clus	tered Ins	tances					
					1	18 ( 3	08)					
					2	5 ( 1	03)					
					3	5 ( 1	(8)					
					4	11 ( 2	28)					
												~
					<						>	
Status												
OK										Log	-	×O

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.

🍰 Weka (	Clusterer	Visualize	: 21:59:57	- SimpleKMea	ans (S	-		×
X: Instance	_number (Nu	m)	×	Y: PSM (Nom)				~
Colour: Clus	ster (Nom)		~	Select Instance				~
Re	Clear	Open	Save	t	itter			
Plot:Studen	t-weka.filters	.supervised	l.attribute.Di	scretize-Rfirst-last-	weka.filte	rs.supe	rvised.a	t
F a iT h i s			<b>xxx</b> xxxxxxx	x0000000	× ×			^
e - CF i r-x000000	000000	00000000	:		100 M	š	i i	
0		24.	5	49	4	i.	5 f	~
Class colour	clust	er0 clus	ter1 clust	ter2 cluster3	cluste	- 4		

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:

🐑 🕀 👷 🐨 🖉			<ul> <li>C Q Search</li> </ul>		\$	ė.	n 1	ŵ	⊜	W.	=
STUDENT PRO	OGRESS	PREDICTOR		None - Student Lago Search Berry	NaT layo	Canada Q	14 				
	HOM:	CLASSIFICATION LIBRARY	ADMINSION ALL	MIN							
Update Student Inf	ormation										
vin:	Life UD										
Enter Labwork Maries :	Enter Wild liber	ek Maries									
Enter Attendance Marks :	aten Strice										
Enter Percentage Marks :	p+roentspe										
Enter Belaxiour Marks :	Dehadosi										
		Submit									
24	A.	Sandar Patel Fustimite of Technolo Audited (VI) City-Mumbal Postcode/72p :	gr Stay T Z Sabau	Ip to Date With What's T in f f	lappening						

### Fig. 15. Shows process of marks Updation

Classified Student in Various class : Class A :

ec/ull-woth.it	n	C 9, secon	26	0 + n	e
		Sue [	hukattoris   Stafftoris   Owlast		
STU.	DENT PROGRESS PREDICTOR	Search	Ilere C		
	FOME CLASSIFICATION LIBRARY	ADMISSION ALUMINI			
Table	for Grade Excellent				
	выше	final percentage	Class		
•	Vinsyak kare	76-0	Cheo A		
а	Vrunda Varvella	85.0000150000	Class A		
4	Dhawal Waghela	66.666666666	Class A		
5	Rabul Madhukar Darada	85.000000000	Class A		
1	Prasannakamar Salanya Bundinkeri	05.000000000/	Clice A		
8	tavish blaget	75.0	Class A		
2	Naha Hamble	75-0	Clare A		
10	Achita Matsaillar	83-000003003	Class A		
- 11	Najara Shinda	R <sub>2 100000000000</sub>	(Taxe A		
10	Sandozh Dovizitta	100.0	Class A		
1/	Provin Sector	03.001113.003	Class A		
18	Askesh Sloth	83 005353530005	Chees A		
	Profile Serverst	81.00000000	Class A		



#### Class B :

-					
	name	final_percentage	Class		
	Janhavi Dhumal	91.0000000007	Class B		
	Glinitukade Allahay	ya.6666666667	Class B		
	3 Bhagyazhri Harvahwar Bhoir	91.6666666667	Claze B		
1	4 Riddhi Jangie	83.493222233	Claco B		
	5 varsharani masal	75-0	Class B		
	6 Parech Patel	75-0	Claro B		
1	9 pooja gapte	75.0	Class B		
1	o NEHA SHIRSHATE	5 <sup>R</sup> 3333333333	Class B		
1	6 swapnali yadav	68.666666667	Class B		
3	p Rostni P.Stende	81.0000000000	Class II		
3	13 milihil daupe	92.6666666667	Clasu B		
1	14 Nikita Divay	ga.6666666667	Class B		
	3 Ritch Warks	83-333333333	Claco B		
1	6 Earthit Gandhi	91.6666666667	Claro B		
1	9 Alatti Vadav	B3-3333333333	Class B		
	o America Kadam	R3.000000000	Clase B		
	n Bhampriya Bhole	26.41	Class B		
4	6 Dipesh Abatate	75.0	Class B		
- 4	7 Gaurav Chandorkar	66.0000066697	Class B		



6	Sarial Rena	yr (86/86/86)	Cleve B	
60	Soluli Germanucia	91 080800807	Ches B	
20	Ajay Vijay Salapel	65.656556567	Class 8	
Table	nr Grade Average			
	and and the second s	final percentage	Class	
•	Shradilla masa	06.000000000	Chas C	
68	sapesh karlass	33333333333	Class C	
és:	mjesh kalam	TRADUCTORS	Close C	
70	Second 1	na nanananan	Class C	
72	survive .	50 500505000	GL 0	
70	Ramsch Shinds	00 000300000	Class C	
26	Baird	12 121121212	Class C	
27	Rated	00 000000000	CL-++ C	
78	adadaal	33.5395357755	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

We place on record and warmly acknowledge the continuous encouragement, invaluable supervision, timely suggestions and inspired guidance offered by our guide Prof. Reeta Koshy, Department of Computer Engineering at Sardar Patel Institute of Technology, Mumbai in bringing the report to a successful competition. We are grateful to Dr. D.R. Kalbande, Head of the Department of Computer Engineering for permitting us to make use of the facilities available in the department to carry out the project successfully. Last but not the least, we express our sincere thanks to all of our friends and our parents who have patiently extended all sorts of help for accomplishing the undertaking.

### 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.