Enhancing Educational Outcomes: Data Mining for Slow Learner Identification in Secondary Education

Fatima Al-Hassan*

College of Engineering, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Education is the Backbone of any country and it is very important to improve the education strength of the country. There are various methods and challenges that the government use these days in schools to improve the standard of education like technologies like computers, smart rooms, projectors, and eBooks. But these recourses are only use full when we know which student need which type of resource or in other words if we predict the results of students, we can improve results and decrease drop out ratios. In this research, we use Data Mining in education to predict patterns based on certain criteria to improve the results of the schools. In this research paper first we explain what data mining is and how it is useful in education and how to predict certain patterns as a result of applying data mining with the use of Weka Tool.

KEYWORDS: KDD (Knowledge Discovery in Database), Education, WEKA data mining tool.

I. INTRODUCTION

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Data mining has been applied in a great number of fields, including retail sales, bioinformatics, and counter-terrorism or even in education. Data mining tools predict future trends and behaviours, allowing institution to make proactive, knowledge-driven decisions. The automated, potential analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer institution questions that traditionally were very time consuming to resolve [5]. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example, predict with 85 percent accuracy which students will or will not graduate. The university could use this information to concentrate academic assistance on those students most at risk.

In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: classification, categorization, estimation, clustering and visualization [1]. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates [2]. Higher education institutions can use classification, for example, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success.

II. DATA MINING TOOLS

- Machine learning
- Computer science, heuristics and
- Induction algorithms
- Artificial intelligence
- Emulating human intelligence
- Neural networks
- Biological models and Engineering

III. PHASES OF DATA MINING

Data mining is an iterative process that typically involves the following phases:

- Problem definition
- Data exploration
- Data preparation
- Modeling

- Evaluation

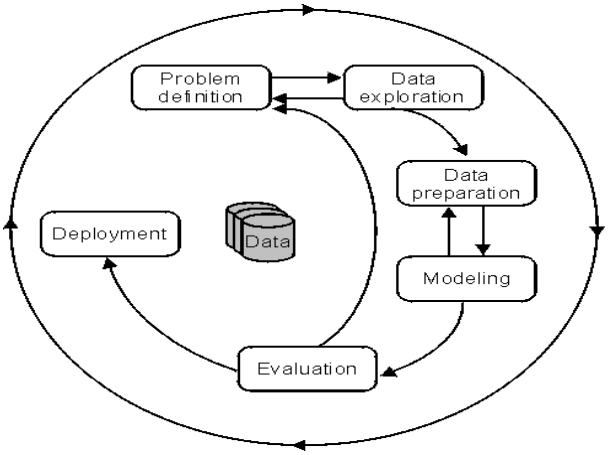


Figure 1 Process of Data Mining

A data mining project starts with the understanding of the problem. Data mining experts, business experts, and domain experts work closely together to define the project objectives and the requirements from a business perspective [4].

In my project our domain is academic data like a student records, result of colleges of different years, strength of students per year and per department and the experts of that domain are HOD's of departments and principals of the colleges.

In the data exploration phase, traditional data analysis tools, for example, statistics, are used to explore the data. In the data preparation phase, data is tweaked multiple times in no prescribed order. Preparing the data for the modeling tool by selecting tables, records, and attributes, are typical tasks in this phase. The meaning of the data is not changed [3].

We select and apply various mining functions because we can use different mining functions for the same type of data mining problem. Some of the mining functions require specific data types.

In the modeling phase, a frequent exchange with the domain experts from the data preparation phase is required. Evaluate the model. If the model does not satisfy their expectations, they go back to the modeling phase and rebuild the model by changing its parameters until optimal values are achieved. When we are finally satisfied with the model, we deployed it.

IV. TOOLS OF DATA COLLECTION & ANALYSIS

Various tools are needed for that project some for analyzing data, some for designing, implementation and some developing software tool these are:

- Excel
- Ms access
- J48 algorithm
- Naïve Bayesian Classifier
- WEKA data mining tool
- Tangara data mining tool
- Rapid miner

V. ADVANTAGE OF DATA MINING IN ACADEMICS

Data mining tells us following things like:

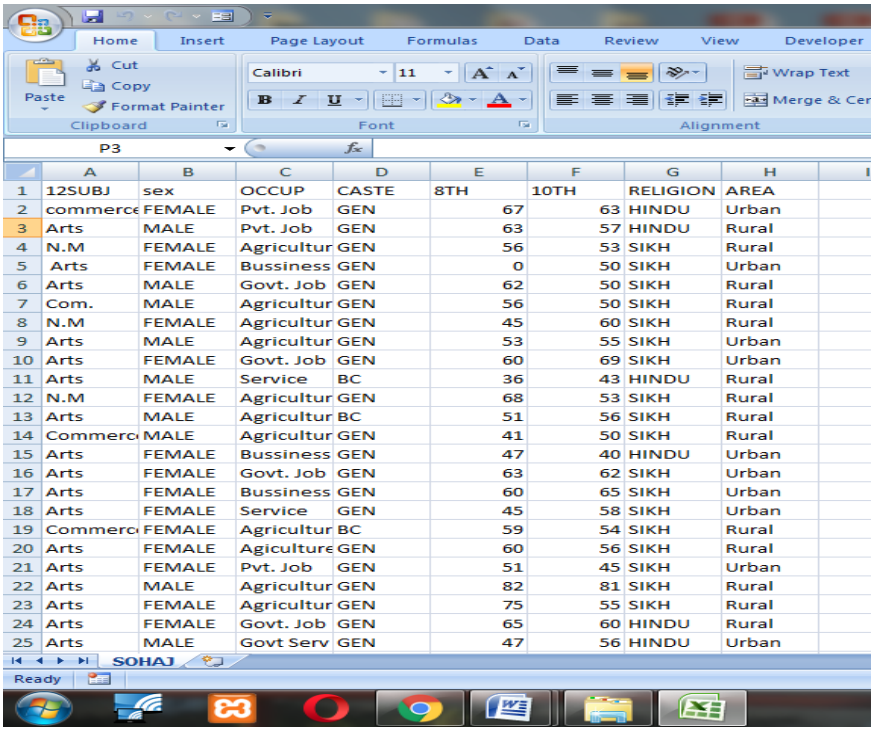
- Tells us about the weak students.
- Tells us which students are taking more credit hours.
- Subjects which are more interesting to students.
- Type of courses we can offer to attract more students.
- Tells various ways to help weak students.
- Helps in improving the result of schools.
- Helps in predicting the result of students.

VI. DATA MINING IMPLEMENTATION

In this research work we collect data of Eight Hundred and thirty eight students of senior secondary schools of Fatehgarh District of Punjab India. In the first step we clean and integrate data. For our problem we chose eight attributes these converted into its equivalent values which are given below in the table.

Table I: Selected Attributes

S.No.	Given Attributes	Description
1	10^{th} percentage	Percentage
2	12^{th} Subjects	Arts , N.M , Commerce , Arts
3	Sex	Male , Female
4	Occupation	Govt , Private , Farmer
5	Caste	Gen , SC , BC
6	8^{th}	Percentage
7	Religion	Sikh, Hindu, Muslim
8	Area	Rural , Urban



	A	B	C	D	E	F	G	H
1	12SUBJ	sex	OCCUP	CASTE	8TH	10TH	RELIGION	AREA
2	commerce	FEMALE	Pvt. Job	GEN	67	63	HINDU	Urban
3	Arts	MALE	Pvt. Job	GEN	63	57	HINDU	Rural
4	N.M	FEMALE	Agricultur	GEN	56	53	SIKH	Rural
5	Arts	FEMALE	Bussiness	GEN	0	50	SIKH	Urban
6	Arts	MALE	Govt. Job	GEN	62	50	SIKH	Rural
7	Com.	MALE	Agricultur	GEN	56	50	SIKH	Rural
8	N.M	FEMALE	Agricultur	GEN	45	60	SIKH	Rural
9	Arts	MALE	Agricultur	GEN	53	55	SIKH	Urban
10	Arts	FEMALE	Govt. Job	GEN	60	69	SIKH	Urban
11	Arts	MALE	Service	BC	36	43	HINDU	Rural
12	N.M	FEMALE	Agricultur	GEN	68	53	SIKH	Rural
13	Arts	MALE	Agricultur	BC	51	56	SIKH	Rural
14	Commerc	MALE	Agricultur	GEN	41	50	SIKH	Rural
15	Arts	FEMALE	Bussiness	GEN	47	40	HINDU	Urban
16	Arts	FEMALE	Govt. Job	GEN	63	62	SIKH	Urban
17	Arts	FEMALE	Bussiness	GEN	60	65	SIKH	Urban
18	Arts	FEMALE	Service	GEN	45	58	SIKH	Urban
19	Commerc	FEMALE	Agricultur	BC	59	54	SIKH	Rural
20	Arts	FEMALE	Agiculture	GEN	60	56	SIKH	Rural
21	Arts	FEMALE	Pvt. Job	GEN	51	45	SIKH	Urban
22	Arts	MALE	Agricultur	GEN	82	81	SIKH	Rural
23	Arts	FEMALE	Agricultur	GEN	75	55	SIKH	Rural
24	Arts	FEMALE	Govt. Job	GEN	65	60	HINDU	Rural
25	Arts	MALE	Govt Serv	GEN	47	56	HINDU	Urban

Fig 2 Csv File of Database

After collecting and cleaning the data we classify data using weka data mining tool. For Classifying and for prediction we use J48 algorithm. For the classification learning experiments the J48 method was chosen (based on the C4.5 algorithm from the machine learning), for being one of the most used Weka classification algorithms that offers a superior stability between precision, speed and interpretability of results [3]. J48 classifies data in the form of decision tree. From this decision tree we are easily identify the weak students and whose chance of failure are maximum and minimum.

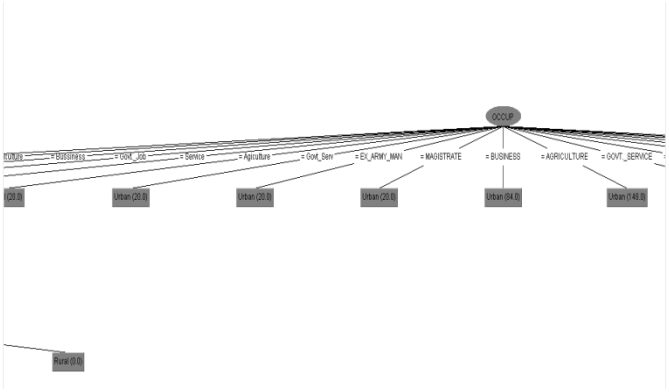


Fig Decision Tree

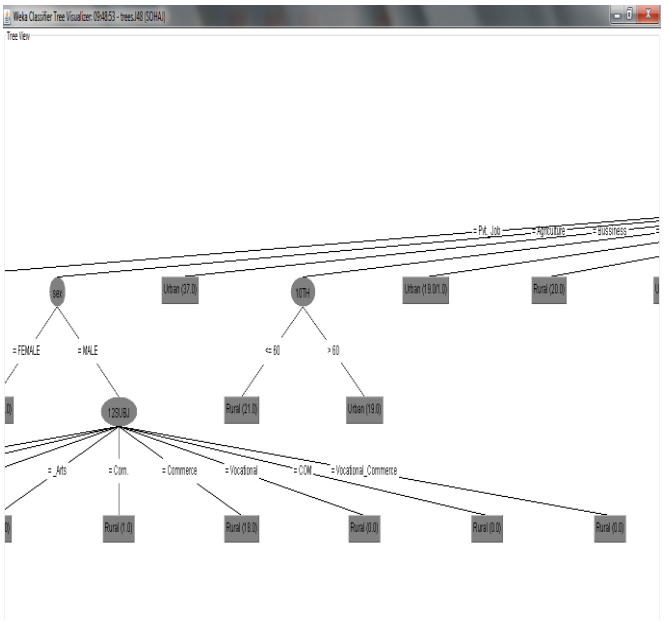


Fig 4 Decision Tree

Statically Results Given by J48 Algorithm

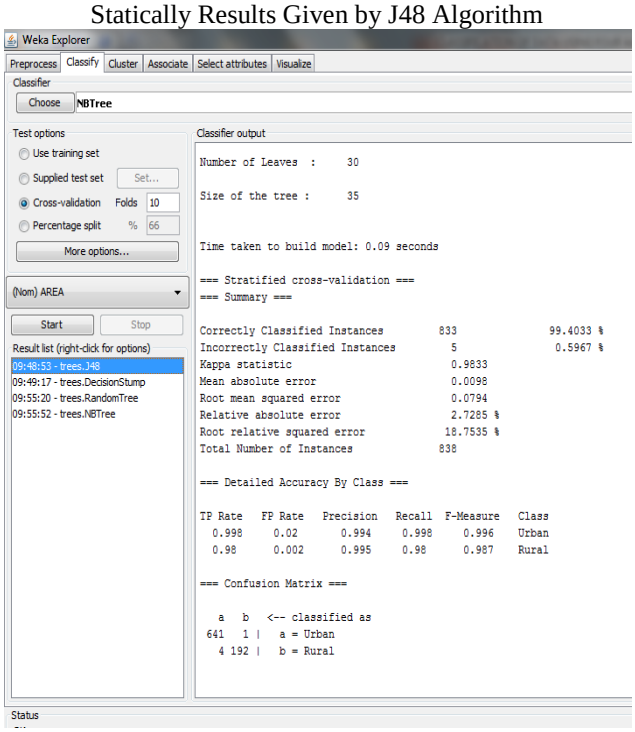


Fig 5 Data Classification Result

The classification learning was also used to predict the students' failure/success to pass the academic exams based on their present behavioural profile. For the J48 classification learning based on the training set, there was 99.4% success rate (the correctly classified instances) which is highest value of prediction. We have 838 instances from which 833 are correctly classified and 5 are not correctly identified.

From the decision tree we are easily identify the weak students and whose chances of fail are maximum. After identifying the weak students we can work hard on that students to minimize the failure result and we can improve overall result and performance of the student.

VII. CONCLUSION

The current education system does not involve any prediction about fail or pass percentage based on the performance. The system doesn't deal with dropouts. There is no efficient method to caution the student about the student about the deficiency in attendance. It doesn't identify the weak student and inform the teacher. Another common problem in larger number of students in class may feel lost in the crowd. Whether they're struggling to find help with coursework, or having difficulty choosing (or getting into) the courses they need, many students are daunted by the task of working through the bureaucracy. Since with the help of Data Mining we identifies the weak students, the teachers can provide academic help for them. It also helps the teacher to act before a student drops or plan for recourse allocation with confidence gained from knowing how many students are likely to pass or fail. Proposed system also shows data graphically according to the need or organization which help them to take important decisions. For future work we also use clustering, with the help of clustering we can see the domain and interest of students in particular field.

VIII. REFERENCES

[1] Hideko Kitahama, "data mining through cluster analysis evaluation on internationalization of universities in japan".

[2] Bruce l. Golden r. H. Smith School of Business University of Maryland, College park, md 20742 "an example of visualization in data mining"

[3] Jing Luan, PHD chief planning and research officer, Cabrillo College founder, knowledge discovery laboratories "Data Mining Applications in Higher Education".

[4] Thulasi, kumarthulasi.kumar@uni.edu, university of northern iowa "theoretical basis for data mining approach to higher education research".

[5] N.V.Anand Kumar Research Scholar, Department of Computer Science and Engineering, Anna University, Chennai "Improving Academic Performance of Students by Applying Data Mining Technique".

[6] Han, J. W., Kamber, M., 2006. Data Mining: Concepts and Techniques, 2nd Edition, The Morgan Kaufmann Series in Data Management Systems, Gray, J. Series Editor, Morgan Kaufmann Publishers.

[7] Luan, J., 2002. Data mining and knowledge management in higher education – potential applications. In Proceedings of AIR Forum, Toronto, Canada.

[8] D. A. Alhammadi and M. S. Aksoy, "Data Mining in Education – An Experimental Study," International Journal of Computer Applications, vol. 62, no. 15, pp. 31-34, 2013. [8]. B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[9] Alaa el-Halees (2009) Mining Students Data to Analyze e-learning Behavior: A Case Study.

[10] U. K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN: 0975-9646, 2011.

[11] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.

[12] Berry, M.J.A., & Linoff, G.S. (1997). Data Mining Techniques for Marketing, Sale, and Customer Support, New York: John Wiley & Sons, Inc.

[13] Chen, S.Y. (1999). Data Mining in Acquiring Association Knowledge, Between Diseases and Medicine Treatments. Unpublished Master's Thesis, National Sun Yat-Sen University, Kaohsiung, Taiwan R.O.C.

[14] Codd, E.F. (1993). Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. E.F. Codd and Associates.

[15] Dai, C.Y. & Sung, J.K. (2005). Data Warehouse of TVC Course in Taiwan, ED-MEDIA 2005, Montreal, Canada, June 27-July 2,2005, AACE.