Use of Data Mining in Education Sector

Manpreet Singh Bhullar, *Member IAENG*, Amritpal Kaur

ABSTRACT— The new interesting subject that offered by institution to interact more student is "DATA MINING". In this paper we will discuss about the problem that are faced by higher education institutes. One of the biggest challenges that higher education faces today is predicting the paths of students. Institutions would like to know, which students will enrol in which course, and which students will need more assistance in particular subject. Also some time management need more information about student like their result, about the success of new offered courses, the answer of all these problems is "Data Mining". Data Mining helps to institution to take decision more accurately. Data mining is better tool to predict the result of the student. In this paper we will discuss about data mining , their different phase's, advantages and also we classify data using weka data mining tool which helps to understand the data. In this paper we use J48 algorithm to predict the result of the student.

Index terms: Induction Algorithm, Knowledge Discovery

I .INTRODUCTION

Data mining is a technique of extraction hidden predictive information from large databases; it is a powerful new technology with great potential to help Universities or institutions to focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing institution to make proactive, knowledge-driven decisions. The automated, prospective 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 too 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, predict with more than 80 percent accuracy (result found) which students will or will not graduate. The university could use this information to concentrate academic assistance on those students most risk of failure.

Manpreet Singh Bhullar is Assistant Professor in Khalsa College Patiala, Punjab India Pin- 147001 (Mobile no : 09814099266; e-mail: manpreetsinghbhullar2010@gmail.com).

Amritpal Kaur is Assistant Professor in Khalsa College Patiala, Punjab, India Pin- 147001 (Mobile no : 09780072525; e-mail: lucky_bhullar2003@yahoo.com).

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, 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 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.

Data mining tools and algorithms

- Machine Learning
- Computer science, heuristics and
- Induction algorithms
- Artificial Intelligence
- Emulating human intelligence
- Neural Networks
- Biological models, psychology and engineering

Phases of Data Mining

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

- Problem definition
- Data exploration
- Data preparation
- Modeling
- Evaluation

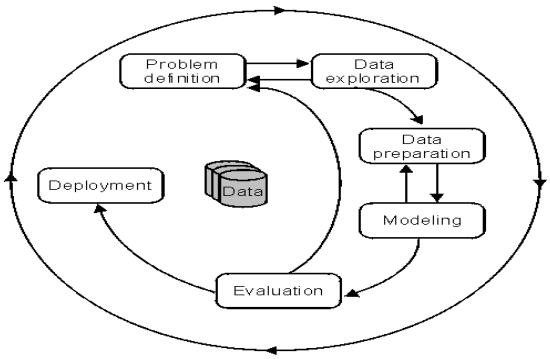


Fig 1 Data Mining Process

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

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 this 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, like 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.

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:

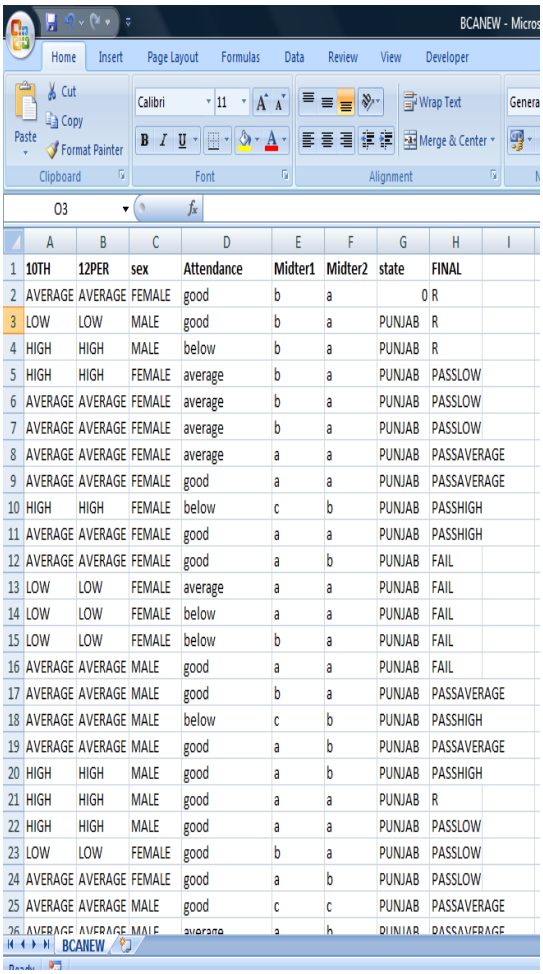
- MYSQL DATBASE
- EXCEL
- MS ACCESS
- SPSS
- METLAB TOOL
- WEKA DATA MINING TOOL
- TANGARA DATA MINING TOOL
- WEB MINER
- V.B 6.0

II. DATA MINING EXPERIMENT

In this research work I collect data of two thousand students from two colleges. 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	Average , Low , High
2	12^{th} percentage	Average , Low , High
3	Sex	Male , Female
4	Attendance	Good , Average, Below
5	Midterm1	A , B , C , D , F
6	Midterm2	A , B , C , D , F
7	State	Punjab, Haryana, Other
8	Final(result of last class)	R , Pass low , Pass average, Pass high , Fail



	A	B	C	D	E	F	G	H
1	10TH	12PER	sex	Attendance	Midter1	Midter2	state	FINAL
2	AVERAGE	AVERAGE	FEMALE	good	b	a	0	R
3	LOW	LOW	MALE	good	b	a	PUNJAB	R
4	HIGH	HIGH	MALE	below	b	a	PUNJAB	R
5	HIGH	HIGH	FEMALE	average	b	a	PUNJAB	PASSLOW
6	AVERAGE	AVERAGE	FEMALE	average	b	a	PUNJAB	PASSLOW
7	AVERAGE	AVERAGE	FEMALE	average	b	a	PUNJAB	PASSLOW
8	AVERAGE	AVERAGE	FEMALE	average	a	a	PUNJAB	PASSAVERAGE
9	AVERAGE	AVERAGE	FEMALE	good	a	a	PUNJAB	PASSAVERAGE
10	HIGH	HIGH	FEMALE	below	c	b	PUNJAB	PASSHIGH
11	AVERAGE	AVERAGE	FEMALE	good	a	a	PUNJAB	PASSHIGH
12	AVERAGE	AVERAGE	FEMALE	good	a	b	PUNJAB	FAIL
13	LOW	LOW	FEMALE	average	a	a	PUNJAB	FAIL
14	LOW	LOW	FEMALE	below	a	a	PUNJAB	FAIL
15	LOW	LOW	FEMALE	below	b	a	PUNJAB	FAIL
16	AVERAGE	AVERAGE	MALE	good	a	a	PUNJAB	FAIL
17	AVERAGE	AVERAGE	MALE	good	b	a	PUNJAB	PASSAVERAGE
18	AVERAGE	AVERAGE	MALE	below	c	b	PUNJAB	PASSHIGH
19	AVERAGE	AVERAGE	MALE	good	a	b	PUNJAB	PASSAVERAGE
20	HIGH	HIGH	MALE	good	a	b	PUNJAB	PASSHIGH
21	HIGH	HIGH	MALE	good	a	a	PUNJAB	R
22	HIGH	HIGH	MALE	good	a	a	PUNJAB	PASSLOW
23	LOW	LOW	FEMALE	good	b	a	PUNJAB	PASSLOW
24	AVERAGE	AVERAGE	FEMALE	good	a	b	PUNJAB	PASSLOW
25	AVERAGE	AVERAGE	MALE	good	c	c	PUNJAB	PASSAVERAGE

Fig 2 Csv File of Database

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

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.

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 a 77.74% success rate (the correctly classified instances) which is highest value of prediction. We have 1892 instances from which 1471 are correctly classified and 421 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.

Advantage of Data mining in Academics

Data mining gives the answers of questions like:

Q: Who is the weak student?

Q: Who are the students taking most credit hours?

Q: The interesting subject of the students?

Q: What type of course can we offer to attract more students?

Q: How can help weak student?

Q: How improve the result of college?

Q: Predicting the result of students?

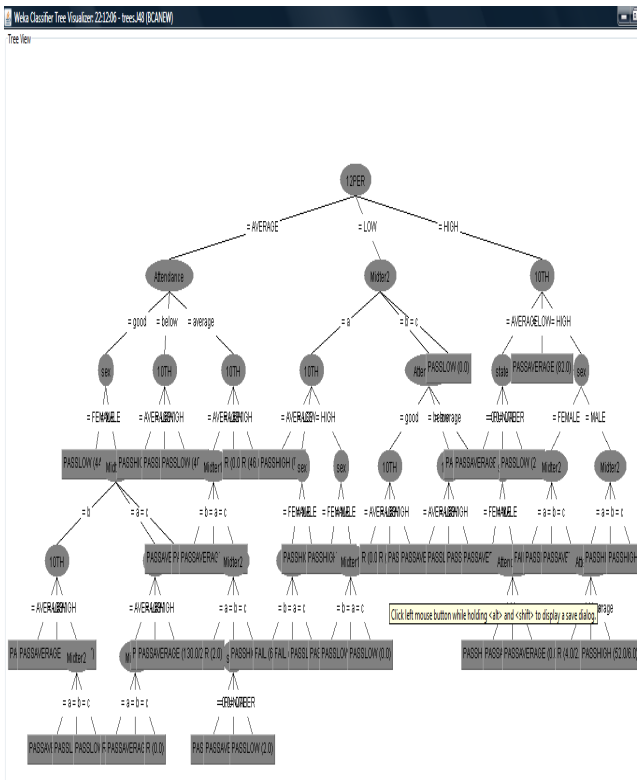


Fig Decision Tree

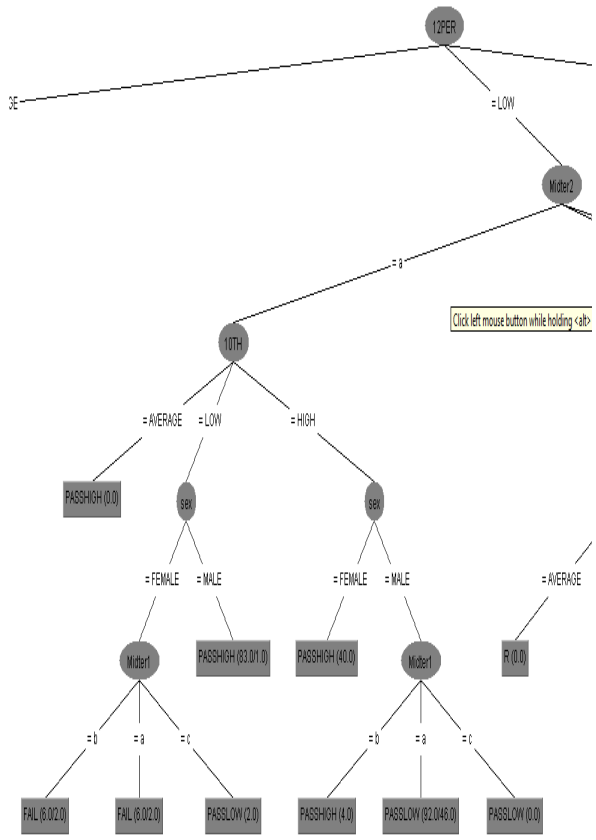


Fig 4 Decision Tree

Statically Results Given by J48 Algorithm

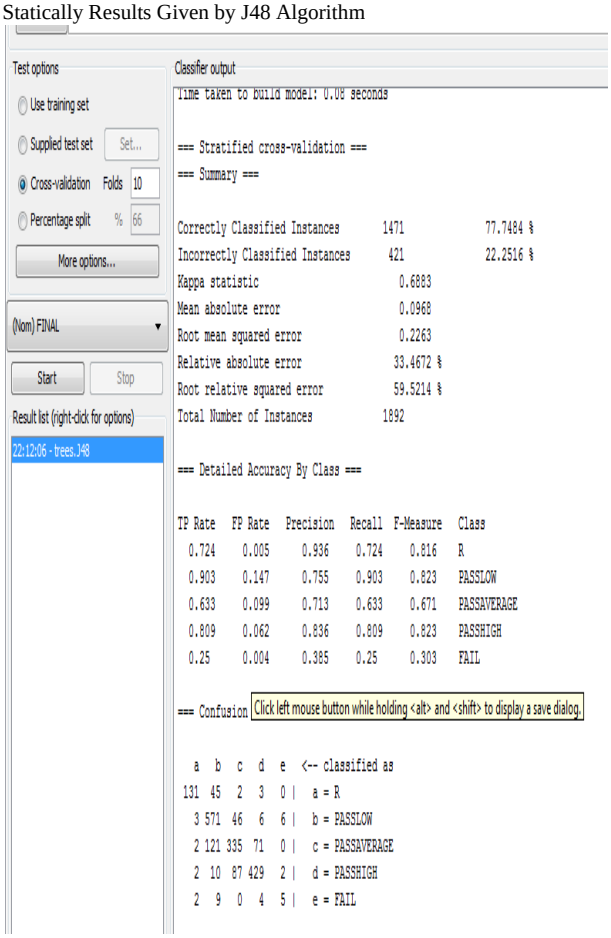


Fig 5 Data Classification Result

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

III. 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 colleges and universities, some students 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 collegiate bureaucracy. Since the proposed model 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.

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 G.V.Uma Assistant professor, Department of Computer Science and Engineering Anna university, Chennai "Improving Academic Performance of Students by Applying Data Mining Technique".

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)