RESEARCH ARTICLE OPEN ACCESS

A STUDY ON THE AWARENESS LEVEL OF EDM & TOOLS

K.Selvi[1], Dr.B.Mukunthan[2]

[1]Research Scholar , Department of Computer Science, Jairams Arts & Science College -Karur

[2]Assistant Professor, Department of Computer Science, Jairams Arts & Science College -Karur

Abstract:

The aim of this research paper is to study the awareness level among the research scholars of various disciplines, educational administrators and the post graduate students regarding the data mining software tools and its applications. It focuses the infrastructure and software facilities and the implementation of educational data mining tools on decision making processes available in the government colleges and the private colleges in a particular university system. This paper surveys the data mining software tools and the awareness among the researchers and post graduate students of different disciplines. First it studies the awareness level among the researchers, and then it analysis the infrastructure facilities available in the university affiliated colleges. This research work show how the data mining software knowledge influences the decision making process. Data mining tools helps the researchers and administrators to arrive a accurate decision through the application of statistics. Educators are possible to predict the learners' performance based on their students' attendance, punctuality, task and project completion capability, mobile usage, infrastructure availability and their personal factors such as family background, peer groups, intelligence, past performance, etc. Those who have learnt more than two data mining tools had better in their research works. Knowing data mining tools enhances the researcher's capability in statistics which leads to more and frequent journal publication related to their research. Computer science graduates and researchers have more knowledge in the data mining & statistics methods than the other discipline researchers**.**

Keywords **— Educational data mining, Opinion about the software tool, Decision making process, Awareness level, Publications, opinion, etc.**

I. INTRODUCTION

Educational data mining is an interdisciplinary research field which concerns education, data mining, statistics and machine learning. Information gathered from colleges and educational institutions has much potential knowledge about the teaching, learning, educational administration, infrastructure facilities, etc. Due to the digitization of vast educational data in an educational set up can be easily processed and visualized to get a fruitful new knowledge by using various algorithms and data mining software tools. There are four main data mining software tools are widely used to process the data in the educational set up. They are WEKA, R programming, SPSS and XL Miner. This type of EDM research helps to predict the student's exam performance to reduce the failure rate. Educational data mining consists of various methodologies used in extracting unknown and hidden useful data by using classifying, grouping and summarizing methods. It gives an idea about the future occurrence. It gives an idea about the descriptive knowledge to find patterns of occurrence.

Data mining methods are classified into association, regression, and classification and clustering. It helps to find the hidden patterns and visualization of grouped data within a fraction of seconds. Association and clustering is called predictive and regression and classification methods are called descriptive. Data mining can be easily performed by using various software tools. Software is useful to handle the huge amount of data. The most widely used data mining tools are WEKA, 'R' programming, SPSS, XL Miner, etc. Learners' preference, learners' attitude, motivation, performance and teaching methodologies applied in an educational set up are converted and analyzed easily by using statistics and data mining tools to predict the accuracy. Data mining tools reduces the manual calculations and difficulties in the huge data handling and calculations

Objectives

- To find the general awareness on educational data mining & tools among the computer science research scholars.
- To find the general awareness on educational data mining & tools among the other discipline research scholars than computer science.
- To analyze the general awareness on educational data mining tools and software among the educational administrators.
- To study the difference if any in educational data mining tools and decision making process among the educational administrators.
- To study the software facilities available in the government and private colleges in a given area and its advantages.
- To study the relationships between tools' knowledge and research publications.

.

2. Related Works

Jiawei Han et al. conducted research related to data mining and its processes. According to his view the processes includes data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. Cristobal Romero et al. pointed out the importance of data mining on effective education. He stressed the importance of different approaches for different educational set up.

Mitra et al. conducted research regarding the categorization of data mining tools based on their hybridizations used, functions implemented and the preference criterion selected, etc. They also studied the data mining issues and problems related to algorithms. Romero surveyed the educational data mining tools and the approaches in a web based learning course environment as well as traditional environment. They found the works needed to find the new knowledge related to educational data mining. Alex conducted research related to interdisciplinary nature of data mining and knowledge discovery in educational data. He pointed the preprocessing and post processing of data, genetic algorithms and evolutional algorithms in his research work. Ligiang et al. surveyed the patterns in data mining and its rankings to get the priority. They studied the interestingness measurements for the data mining. They also studied the future development of pattern and measurement.

Shiguo wang conducted a survey on algorithms, data mining, statistical test, performance, etc. They discussed about the decision tree, approaches, accuracy of models in data mining. They suggested suitable algorithm to develop the accuracy. Michael conducted survey about the emerging data mining tools and studied about its accuracy in prediction. He classified the software based on their characteristics, database connectivity and data mining characteristics. Milovic and Boris conducted study on health care data mining to predict the patience behavior, and future occurrence of communicable deceases. They studied the hidden patterns to find out the relationships. They used the strategy in decision making process in the health care industry.

Jing Luan conducted the data mining applications and its impact on higher education. He found out he important of data mining tools in the educational set ups. Brijesh Kumar et al. applied the data mining tools in the educational data to find out the performance of the learners. They successfully implemented the data mining software application to predict the decision making process to find out the performance. They studied the classification method to find out the student performance in the end semester examination and the decision tree method.

Picciano and Anthony studied the big data and learning analytics on higher education. They pointed the details of learning analytics and the supportive processes for educational administrators. They conducted the survey about the big data tools, concepts, applications, growth areas in the American higher education system. Carlo studied the data mining processes related to decision making support for business. He extensively studied the decision support system for the business and commerce industry. Business

intelligence system and its usage to forecasting the ups and downs were discussed by him in his book named 'Business intelligence'.

3. Hypothesis

(i) There is a significant difference in educational data mining software knowledge among the computer science and other discipline researchers.

(ii) There is a significant difference in awareness level among the educational administrators among the private and government institutions.

(iii) There is a significant difference in educational data mining tools and software usage among the government and private institutions.

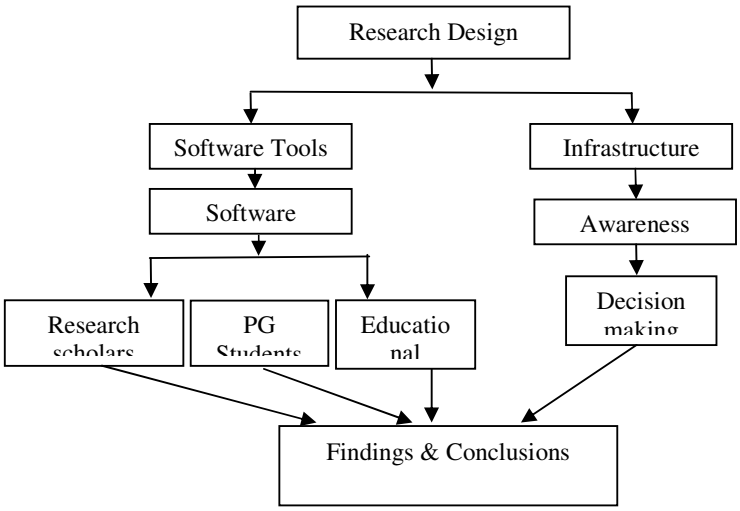
(iv) There is a relationship between the awareness level and the decision making process among the educational administrators.

(v) There is a strong difference between the data mining research scholars of computer science researchers versus other researchers related to computer science in software tools practical knowledge and statistics.

(vi) There is a significant difference among the post graduate students and research scholars in educational data mining related knowledge and tools awareness.

(vii) There is a significant difference among the researchers in data mining tools knowledge and research paper publications

(viii) Those who have more data mining tools knowledge had more published research papers.



a. Hypothesis Methodology

3.1 Sampling

Convenient sampling technique was used to collect the data. Total 230 research scholars from computer science, post graduate students from computer science and other discipline participated in the research. Total 120 educational administrators from private and government colleges participated and answered voluntarily. The secretaries, principals, directors and head of the staffs included in the educational administrator's category. A questionnaire constructed with the careful consideration and opinion of validity and reliability. The questionnaire pre tested with the help of other experienced researchers.

Reliability:

If the result of the measurement has similarity it is said to be reliable. Under the same condition the results and findings should have similarities. High reliable results are possible through reliability coefficient calculations. It ranges from .00 to 1.00 which indicates the errors and possibility of confidence levels.

Parallel-forms method:

Two parallel content forms are used to collect the data to get the reliable data from the respondents. This type of psychometric approach is helpful to reduce the error. Two questionnaire with the same content, response items, statistical characteristics are given to find out the knowledge about the software like, WEKA, R programming, XL miner, SPSS. Apart from the opinion

questionnaire, the infrastructure related to software and facilities are also tested. The subject and software knowledge is tested by using two different questionnaires to find the reliable results. The correlation between form A and B gives the reliable results.

4 Findings of the study

1) There is a significant and positive relationship between educational data mining knowledge and the software tools usage among the computer science research scholars
2) There is a significant difference exists between the computer science researchers and other discipline researchers in tools usage and basic statistics.
3) There is a significant relationships exist among the educational administrators in private and government institutions.
4) There is a relationships exists between the decision making process and tools awareness level and application knowledge.
5) The awareness level among the educational administrators regarding the data mining tools and infrastructure requirements in very low.
6) More than 80 % of Post graduate students have no knowledge about the term learner analytics and impact study.
7) There is a difference exists between the post graduate students of various discipline.
8) More than 90 % research scholars' uses at least one software tool.
9) 70 % of the computer science post graduate students have no knowledge about the practical application of the data mining tool.
10) 20 % private institutions have licensed version of statistical tools.
11) More than 60 % government colleges have open access data mining tools for their research purpose.
12) 40 % of the researchers have knowledge about the algorithms related to data mining
13) 90 % of the computer science researchers opined favorably the concept of algorithm improvements

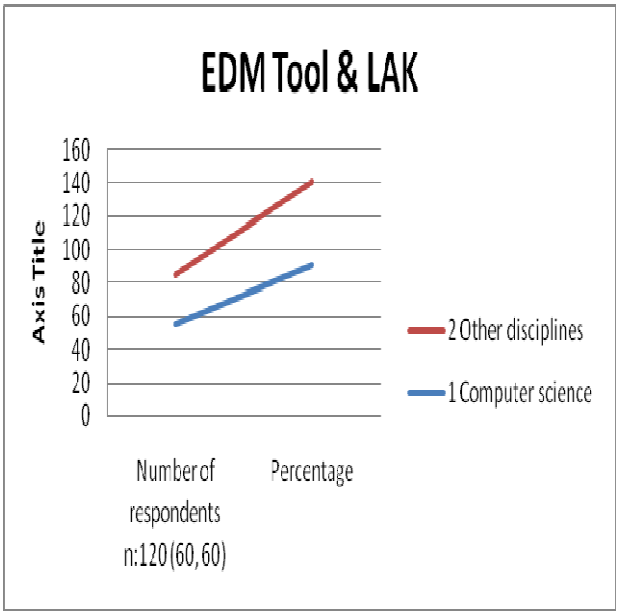
Tables

The questionnaire given to collect the data had six basic parts with 30 questions

They are listed below

S.No	Contents	Further classification
1	Software awareness	Types, usage, input, output, User interface, etc.
2	Basic concepts and knowledge	EDM, LAK, impact study
3	Infrastructure	Software, licensed, open source, lab facility, etc.
4	Decision making process	EDM knowledge & decision making steps
5	Statistics and tools application	Basic idea, implementation
6	Opinion of the respondents	Satisfaction, flexibility, usage , etc.

Questionnaire	Responses 1 to 5 points Strongly Agree to Strongly disagree
1. The software is flexible to use	
2. Visualization & graphical representation is simple	
3. Easy statistical input and formula	
4. Easy for me to learn and apply	
It is compatible with other software	



Profile of the Respondents:

- All the respondents (students and research scholars) are the degree holder with the age groups between 24 to 34
- All the educational administrators are above the age group of 40 years.
- 90 % of all the respondents had known the basic idea about data mining.
- 90 % colleges of government and private colleges uses open source software
- 20 % private colleges uses licensed version of DM tool.

60 % government colleges do not have adequate software facility and tools to process huge data

Table 1
Respondents' opinion about EDM and Learner Analytics & Knowledge

S. No	Knowledge	Number of respondents n:120 (60, 60)	Percentage
1	Computer science	55	91
2	Other disciplines	30	50

Table 2
Data mining tools related knowledge and decision making process-opinion
It will help to get an accurate decision

S. No	User opinion	Number of respondents n:120	Percentage
1	Yes	95	79
2	No	35	21

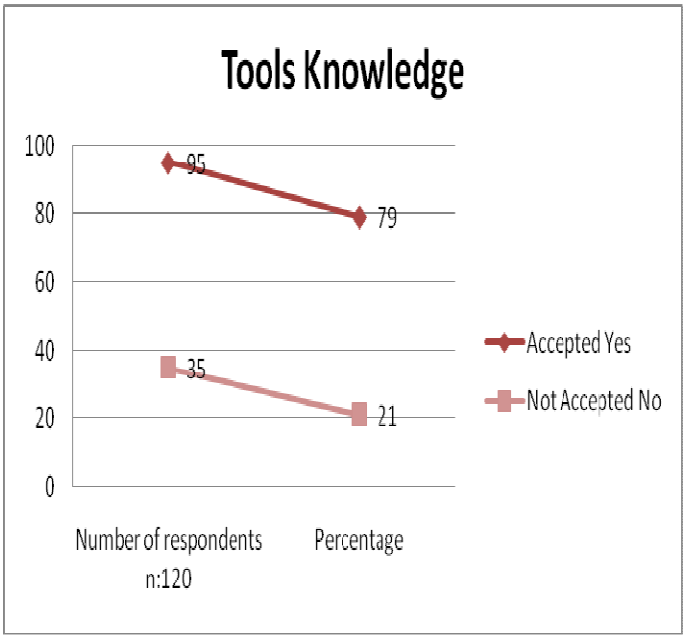


Table 3

Researchers' opinion about the Data mining tools Flexible and easy to apply

S. No	User opinion	Number of respondents n:230	Percentage
1	Yes	190	82
2	No	40	18

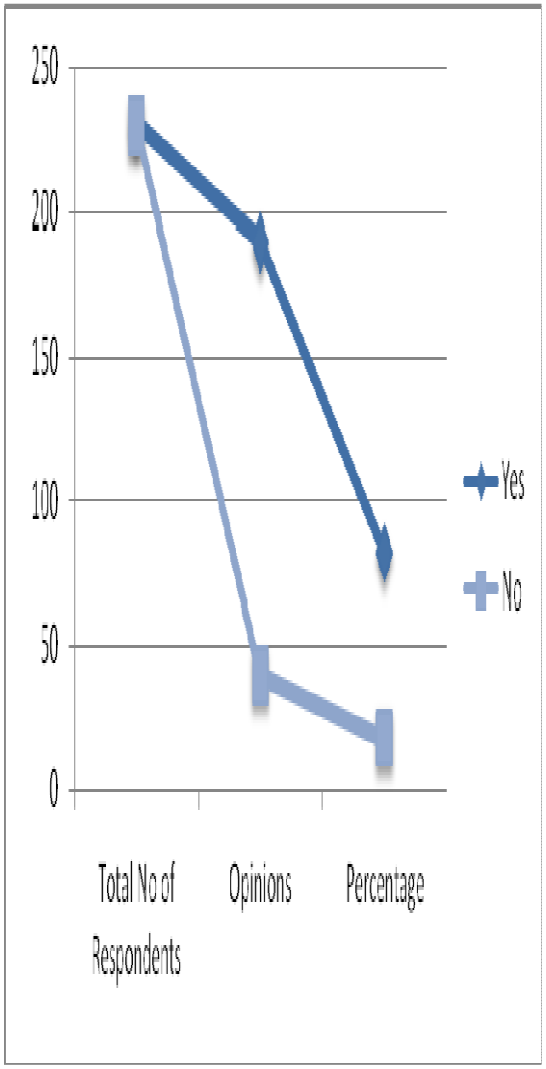
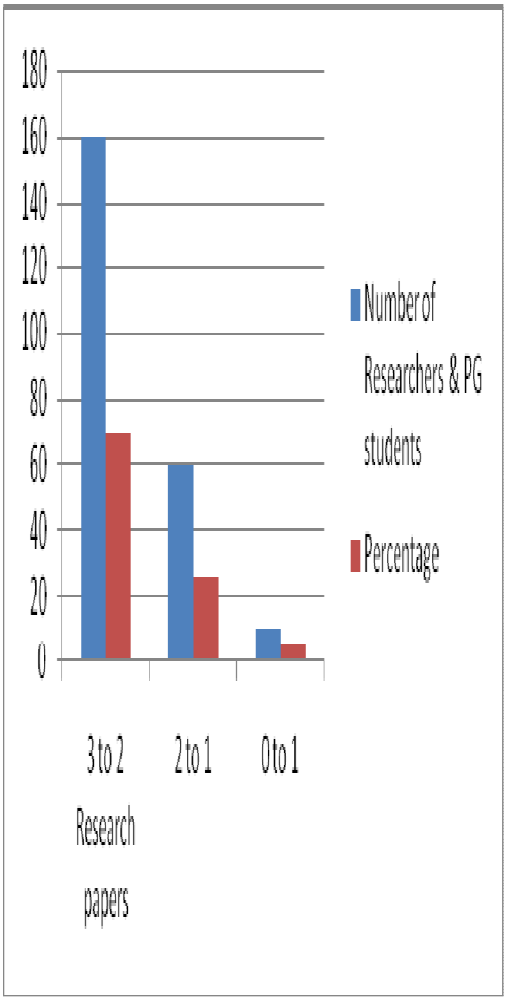


Table 4

Researchers' practical knowledge and research publications

S. No	Research papers	Number of Researchers & PG students 230	Percentage
1	2-3	160	69
2	1-2	60	26
3	0-1	10	5



5. Limitations

Only colleges in and around Trichy district is considered for the research. Other discipline researchers from Biology, Physics, Chemistry, Commerce and Language have no knowledge about the data science and tools due to the syllabus. The administrators from various institutions belong to different disciplines. Educational data mining and decision support process is very new to them. This research considers the arts and science discipline institutions in Trichy district. The awareness level cannot be compared to other developed cities. Even though research publication by the researchers are compared with data mining and statistical tools knowledge, many other factors such as infrastructure facility, tools availability, economical aspects should be considered for the better result.

5. Conclusions

This research work concludes that data mining software tools and infrastructure is not adequate both in private and government college. Many researchers opined that educational data mining has a potential and a new and innovative field to do research. Sixty percent of them do not posses knowledge about the accuracy of the software. Even though many of them accept the importance they have no idea about the decision making process and software application. Many of them accept the importance of learner analytics and software to improvements and algorithm improvements are necessary to predict the future occurrence with very minimum errors. Software developers should consider the opinion of the research community. The study about the learners' opinion is one of the important aspect to upgrade the version. The budding computer science researchers have many ideas related to the improvements of the software as well as the algorithms and methods. It should be taken into account to develop a suitable enhancement.. Algorithm related information have to be consolidated in order to develop more accuracy.

6. References:

1. Jiawai Han, Micheline Kamber, Jian Pei, "Data mining concepts and techniques", Elsevier, Edition 3 p 3-6, 2012
2. Cristobal Romero, Sebastian Ventura, "Data mining on education", Wiley online library, pp 1, 2012
3. S. Mitra, S.K Pal. K. Mitra, "Data mining in soft framework- A survey", IEEE transaction, Neural networks, Vol. 13, Issue 1, pp 3-14, 2002
4. C. Romero, Sebastian ventura, "Educational data mining – A Survey from 1995-2005, Science direct, Vol, 33 issue 1, pp.135-146, 2007
5. Alex A. Freitas, "A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery", Natural computing book series, NCS, Advances in Evolutionary Computing pp 819-845
6. Liqiang, Howard J. Hamilton, "Interestingness measures for data mining: A survey", ACM Computing Surveys (CSUR), Volume 38 Issue 3, 2006
7. Shiguo Wang, "A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research", IEEE, Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference, 2010.
8. Michael Goebel, Le Gruenwald "A survey of data mining and knowledge discovery software tools", ACM SIGKDD Explorations, Volume 1 Issue 1, pp.20-33. June 1999.
9. Milovic , Boris, "Prediction and Decision Making in Health Care Using Data Mining", Kuwait Chapter of Arabian Journal of Business and Management Review, American University, 2012.
10. Jing Luan, "Data Mining and Its Applications in Higher Education", Wiley online library, 2012.

11. Brijesh Kumar Baradwaj, Saurabh Pal "Mining Educational Data to Analyze Students' Performance**",** Cornell university publications, IJACSA, 2012

12. Picciano, Anthony G, "The Evolution of Big Data and Learning Analytics in American Higher Education" Journal of Asynchronous Learning Networks, San Francisco, CA.: Morgan Kaufmann Publishers, Inc.Vol.16 n3 p9-20 Jun 2012.

13. Callo Vercellis, "Business intelligence", Data Mining and Optimization for Decision Making, John Wiley and sons, pp 2-15, 2009