Application of Educational Data mining Techniques in E-Learning Systems with its Security Issues: A Case Study

Ms. Ankita Chopra[1], Ms. Aakanksha Chopra[2]

Assistant Professor, IT, Jagan Institute of Management Studies, Delhi, India[1, 2]

Abstract: Recently, Educational Data Mining has become an emerging research field used to extract knowledge and discover patterns from E-Learning systems. This work is a survey of the specific application of data mining in learning management systems and a case study with university database.E-learning systems is an information superhighway where large scale communication network is provided with variety of interactive services are provided as text-database, e-mail, audio, video etc. Cyberspace is the indefinite areafamously known as Internet. Cyber security is the frame of protocols framed in one place for safeguarding this cyberspace.This paper explains how data mining tasks like clustering can be applied to the data taken from an e-learning system and also deals with E-threats and E-risks associated with it.This paper will also focus upon ubiquity of internal cyber-attack as well as lack of proper IT policies and procedures in e-Learning systems. The performance of students on online course in digital electronics is taken for the analysis and results are achieved with WEKA tool.

Keywords: EDM, Classification, Clustering, WEKA, E- Threats and E-Risk.

I. INTRODUCTION

Across a wide variety of fields data are being collected and accumulated at dramatic pace. There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. These theories and tools are subject for the merging field of knowledge discovery in databases (KDD).

KDD refers to the overall process of discovery useful knowledge from data, and data mining refers to a particular step in this process. Data Mining is the application of specific algorithms for extracting patterns from data. The widespread application areas of data mining include bio-informatics, ecommerce, fraud detection and now in the field of education as well. The data mining in the field of educational research is known as Educational Data Mining (EDM).EDM tries to simulate a student model which may be used for the improvement of students by predicting the future trends.

The sharing of information, collaboration and interconnectivity are the core elements of any e-learning system. Since e-learning system is open, distributed and interconnected then, the security becomes an important challenge in order to ensure that interested and authorised actors only have access to the right information at the appropriate time [1].

Protection against data manipulation, fraudulent user authentication and compromises in confidentiality are important security issues in e-learning. E-risk involve the risk at the time of electronic transaction, whereas threat means an anticipated danger [2].But in order to have a secure transmission of the information being exchanged over internet, one needs the concept of Network Security,

which needs to take punitive action to Ease of Use protect from different types of attackers like- hackers, interested computer neophytes, deceitful vendors or disenchanted employees of an organization [3]. Common threats for computers are viruses, network penetrations, theft and unauthorised modification of data, eavesdropping and non-availability of servers and personal computers [4].

The challenge is how to elicitate, manage and organize this big data. The actual thing is to find out which data is important, what to keep and what to discard [5].

This paper explains the methods of EDM through an example data taken from an e-learning platform and also deals with its security threats and risks.This paper is organized as: EDM and its important methods like classification,clustering. It also explains how WEKA tool is used to analyse the student data in online digital electronics course taken from university database.What are the various security threats and risks associated with E-Learning systems?
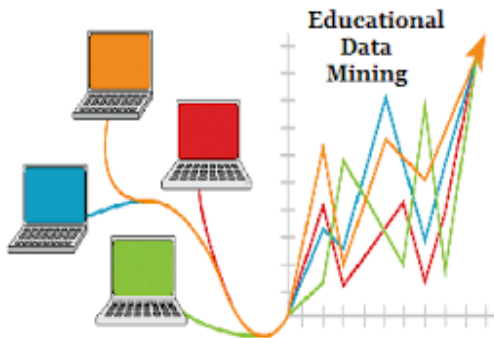
II. EDUCATIONAL DATA MINING

Education is increasingly occurring online or in educational software, resulting in an explosion of data that can be used to improve educational effectiveness and support basic research on learning.

E-Learning can be defined as, the use of internet technologies to deliver a broad array of solutions that enhance knowledge and performance. E-learning is based on three fundamental criteria's they are

- Networked for instant updating distribution, storage/retrieval and sharing of information.

- Content delivery via computer using WWW
- It focuses on the broadest view of learning and learning solutions.

The major challenge in E-learning systems is the collection of unstructured information that chokes the educational system without providing any articulate knowledge to its actor's data mining was born to tackle problems like this.

EDM is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students and the settings which they learn in. EDM develops methods and applies techniques from statistics, machine learning and data

Mining to analyse data collected during teaching and learning.Data Collected from online learning systems can be aggregated over large number of students and can contain many variables that data mining algorithms can explore for model building.

Goals of EDM:
- Predicting student's future learning behaviour by creating student models that incorporate such detailed information as students' knowledge, motivation, metacognition and attitudes.
- Discovering or improving domain models that characterize the content to be learned and optimal instructional sequences.
- Studying the effects of different kinds of pedagogical support that can be provided by learning software and
- Advancing scientific knowledge about learning and learners through building computational models that incorporate models of the student, the domain and the software's pedagogy.

III.WEKA AS EDM TOOL

For the purposes of this study, we select WEKA (Waikato environment for knowledge analysis) software that was developed at the University of Waikato in New Zealand.

WEKA is open source software issued under the GNU general public license. It contains tools for data pre-processing, classification, regression, clustering, association rules and visualization.

It is used in many application areas in particular for education & research.

Data Set used:

This paper considers Digital Electronics course as an E learning environment for a university abroad. The students learn the problem through examples, understand through doing exercises and test their knowledge through tests.

In this e learning environment there are 6 sessions and 100 students are made to participate in each and every session. Tests are conducted after every session to test students' knowledge and analyse their behaviour.

- Data Preprocessing: Total 53 samples were collected at the end of session 1 exam and preprocessing techniques have been applied to these samples to yield the following results

Table I

Student ID	Marks (2)
3	2
6	2
7	2
10	2
13	2
15	2
16	2
17	2
18	1
20	2
24	2
27	2
28	2
29	2
30	0
32	2
36	2
37	2
39	2
41	2
44	2
46	2
48	2
51	1
53	1
56	2
59	2
62	2
63	2
66	2
67	2
68	2
72	2
73	2
74	1
76	2
78	2
81	2
83	2
85	2
86	2

87	2
88	2
89	2
92	1
93	2
96	2
98	2
99	1
100	2
101	2
76	2

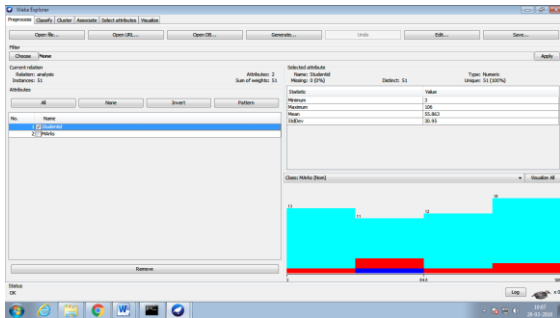After Preprocessing the samples were reduced to 51

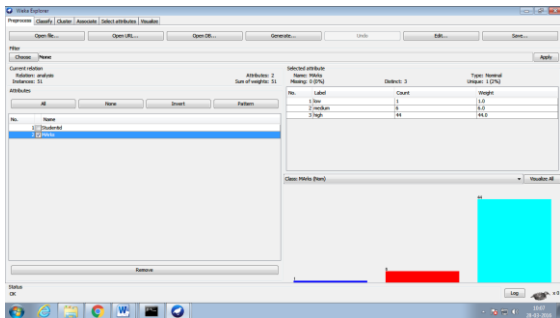Fig1; / (Preprocessing of student id)

Fig2; / (Preprocessing of student marks)

- Application of Data Mining Algorithm :Clustering(K-Means)

The student attributes can be categorized as
- Session Id
- Student Id:{7,10,15,17,20}
- Exercise:{0,1}
- Activity:{0,1,2,3,4,5}

The above Attributes of students are defined as follows:

- Session Id: There are total 100 students participating in 6 sessions in online digital electronics course. Every session has a corresponding session Id
- Student Id: Every student is given a unique ID to identify and evaluate a student correctly from a batch of 100 students/session
- Exercise(Nominal Attribute): It indicates exercise no of a particular session
- Activity(Nominal Attribute):It indicates student is doing one of the following activities
- Study_Es{0}:It indicates that a student is viewing the contect of specific exercise

- Other{1}:Student is not viewing any page
- Deeds_ES{2}: It indicates student is working on a specific exercise.

Study Material{3}: It indicates student is viewing the content.
- Deeds{4}:It indicayes student is not clear what exercise he is working on

Sample Data for session 1 in arff format is as follows:
@Relation analysis
@ATTRIBUTE sessionid NUMERIC
@ATTRIBUTE studentid NUMERIC
@ATTRIBUTE excercise {0,1,2}
@ATTRIBUTE activity {0,1,2,3,4,5}
@Data
1, 7, 0, 0
1, 7,1, 1
1, 7, 1, 2
1, 7,1, 3
1, 7,1, 3
1, 7,1, 3
1, 7,1, 3
1, 7,1, 3
1, 7,1, 1
1, 7, 1, 3
1, 10, 0, 0
1, 10, 0, 0
1, 10, 1, 3
1, 10, 1, 0
1, 10,1, 3
1, 10,1, 0
1, 10,1, 0
1, 10,1, 0
1, 10,1, 0
1, 10,1, 0
1, 15, 0, 0
1, 15, 0, 4
1, 15, 1, 1
1, 15, 1, 1
1, 15, 1, 1
1, 15, 1, 3
1, 15,1, 0
1, 15, 1, 3
1, 15, 1, 1
1, 17, 0, 0
1, 17, 0, 5
1, 17, 0, 0
1, 17, 0, 0
1, 17, 0, 5
1, 17, 0, 5
1, 17, 0, 5
1, 17, 0, 5
1, 17, 1, 1
1, 17, 1, 3
1, 20, 0, 0
1, 20, 0, 4
1, 20, 0, 2
1, 20, 1, 5
1, 20, 1, 5
1, 20, 1, 0

1, 20, 1, 5
1, 20, 1, 1
1, 20, 1, 0

After applying EM-Clustering Algorithm the results by WEKA tool are
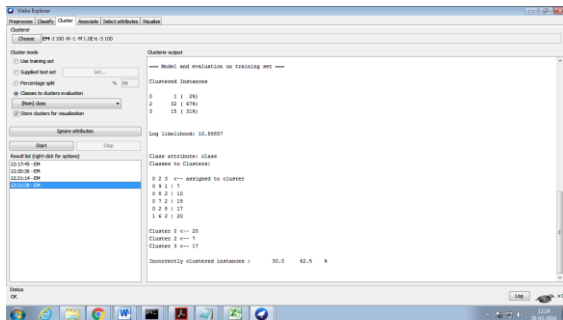
Which clearly shows that
Cluster 0:student id :20
Cluster1:student id:7
Cluster 2:student id:17
Hence, we can analyse the student behaviour by looking at table 1 that a student performs well in exam
i) if the content of the material which he is viewing is excellent.
ii) if he completely understands the exercise of the session which he is doing.
iii) if student gets less involved in doing other things or being idle.

IV.CYBER SECURITY THREATS AND RISKS IN E LEARNING SYSTEMS

Security of Network these days is becoming a very critical facet when the degree of data to be shared is huge and complex. So the important issue today is the information secrecy for the secure data communication over the network; protecting data from being disclosed to illegitimate users while legitimate users are sharing the data. Cryptography in Networks helps for the same as itenhances the data confidentiality and privacy by making information incoherent. Hence, the data or the information stays protected from the illegitimate users.

Data Encryption and decryption are the two most important techniques used today for enhancing the security of the data. **Cryptography** is an art of converting the data on the applications into incoherent or scrambled or in unintelligible format. It related to the study of mathematical algorithms related to aspects of information security such as **confidentiality, data integrity, and data authentication [6].**Symmetric Key Encryption and Asymmetric Key Encryption are other two important encryption types [7].

Building trust and encouraging engagement amongst users of e-learning systems is important and becoming crucial because there are opportunities for both synchronous and asynchronous interactions with the system. Synchronous learning occurs in real-time, with all participants interacting at the same time, while asynchronous learning is self-paced and allows participants to engage in the exchange of ideas or information without the dependency of other participants′ involvement at the same time [1].

A. **E-threats-** A threat in e-learning may cause e-risk. There are major threats such as -Confidentiality violation, Integrity Violation, Denial of service[1,4], Illegitimate Use [6,7], Malicious Program, Repudiation, Masquerade, Traffic Analysis, Brute-force attack, [1,4], Guessing passwords, hijacking, Sniffers, Social Engineering, Spoofing [4]. All these threats are to be dealt individually by stopping the penetration and loop holes.

Due to above mentioned threats following risks may occur. Before even starting with any e-learning system e-risks should be given highest priority.

B. **E-risks**- Institution should perform a cyber security risk assessment. There are many e-risks like author's risk, Teacher's risk, Manager's risk, System developer's risk, Student, risk.

i. **Author's risk**- An author is someone who is writing and developing the content in an e-learning system. Now the major risk at author's end is that there content and notes might get altered or processed without their knowledge. There content might be used in an unauthorized manner, modified and reused in different context in e-learning.

ii. **Teacher's risk**- teacher in an e-learning system is someone who delivers a lecture and use material by different author's. Risk with teacher's is how to deliver a lecture, sending notes and assignments, accepting and marking answer sheets, preparing and distributing mark sheets.

iii. **Manager's risk**- There are three types of risk at managerial level. Firstly, inelegant people dissimulating as students and writing the paper instead of original students. Secondly, maintaining passwords of all servers and routers, recording of daily traffic and looking after other network devices. Thirdly, to control authorization i.e. access strategies.

iv. **System Developer's risk**- there is two types of risk involved as developers end. Firstly, providing quality hardware components like high ended servers for designing, developing, and delivering E-learning products. Secondly, storing passwords in clear text in the application code so that is not accessed by intelligent learner.

v. **Student's risk**- receiving un- altered question paper, should be aware of phishing, and store login information in order to prevent from unauthorised access.

V. CONCLUSION

In this work we have shown how useful Data Mining Techniques can be if applied to Learning Management Systems. They help in analyzing /predicting student behavior pattern and update E-learning systems accordingly to benefit students, instructor, system developer etc. With E-learning systems spreading its

wings in the education sector to reach maximum people security of such systems should be set to the top priority. Further the level of security could be secured by using various cryptographic techniques, DRM, biometric authentications, access control using firewalls etc.

ACKNOWLEDGMENT

REFERENCES

[1] I. Bandra, F. Ioras, K. Maher, "Cyber Security Concerns in E-Learning Education," International Conference of Education, Research and Innovation Seville, Spain. Proceedings of ICERI2014 Conference, Nov 17-19, 2014, Seville, Spain, ISBN: 978-84-617-2484-0, ISSN: 2340-1095, Published: IATED, pp. 0728-0734. [Online], Available: http://ecesm.net/sites/default/files/ICERI_2014.pdf

[2] N. Barik, Dr. S. Karforma, "Risks and Remedies in E-learning System," International Journal of Network Security & Its Applications (IJNSA), vol. 4, No. 1, pp. 51-59, Jan 2012. DOI: 10.5121/ijnsa.2012.4105. [Online], Available:http://arxiv.org/ftp/arxiv/papers/1205/1205.2711.pdf

[3] Aakanksha Chopra, "Security Issues of Firewall," International Journal of P2P Network Trends and Technology (IJPTT)- vol. 22, no. 1, Jan 2016, ISSN: 2249-2615, pp. 4-9. [Online], Available: http://www.ijettjournal.org

[4] Zeng-gang, X., &Xue-min, Z. (2010). Research and Design on distributed Firewall based on LAN. Computer and Automation Engineering (ICCAE), 2010, E-ISBN: 978-1-4244-5586-7, Print ISBN: 978-1-4244-5585-0, INSPEC Accession Number: 11259785, DOI: 10.1109/ICCAE.2010.5451596,pp: 517-520, Publisher: IEEE, Singapore.

[5] Suman Madan, Aakanksha Chopra, "Data Driven Analytics: Prospects and Challenges," Advances in Comuter Science and information Technology (ACSIT), Print ISSN: 2393-9907, Online ISSN: 2393-9915, vol. 2, no. 4, April-June, 2015.

[6] Aakanksha Chopra, Manish Singh, "Key Exchange Algorithms in Cryptography: A roadmap to Data Security," International Conference organized by JIMS

[7] Aakanksha Chopra, "Comparative Analysis of Key Exchange Algorithms In Cryptography and its Implementation," Journals of Innovations- IMS Noida, vol. VIII, issue-2, Dec 2013, Print ISSN: 0974-7141, Online ISSN: 0976-1713.

[8] Cristo´bal Romero *, Sebastia´n Ventura, Enrique Garcia"Data mining in course management systems: Model case study and tutorial, Department of Computer Sciences and Numerical Analysis, University of Co´rdoba, 14071Co´rdoba, Spain Nov17-19, 2014, Seville, Spain, [Online], Available: ww.elsevier.com/locate/compedu

[9] Ankita Chopra," Quantitative Analysis of Dairy Product packaging with the application of data mining techniques," International Journal of Computer science and Information Technology, ISSN: 0975-9646, vol 7, no 2, Mar-Apr, 2016.