Identification of Influencing Factors for Enhancing Online Learning Usage Model: Evidence from an Indian University

Sachin Ahuja a*, Puninder Kaur a, S N Panda a

a Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Received: 18 October 2018; Accepted: 15 February 2019; Published: 08 March 2019

Abstract

With advent of technology online education has become the core of educational settings worldwide. This paper aims to identify the factors that contribute in enhancing the online learning usage model in context of India, an emerging leader in Educational settings across the globe. In this study, data mining techniques were applied on the data collected from the log files of online courses. The initial investigations supported the use of custom build framework for teaching online courses. Data Structure course was taught using online platform and the data was collected using the log files. The data collected was further analysed using data mining techniques using Rapid miner tool. Although the results from three different data mining techniques showed some variations but the inferences from the results identified few common factors that have influence on enhancing the online learning usage model. Clustering techniques revealed that factors related to timely checking of online contents and posting have positive impact on online learning however decision trees supported that timely completing the online assignments along with checking of online contents and posting of messages played an important role in terms of enhancing the academic performance. This paper identifies three factors for teachers teaching online courses to improve overall performance of the students by learning from Indian University Success.

Index Terms: Online Course, Online Learning, Data Mining, Academic Performance.

© 2019 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Data mining has played a pivotal role in the field of Educational data mining. Due to globalization the world has turned into a compact market with utter disregard to physical boundaries, distances and barriers. Online
learning is synergized due to worldwide expansion and advent of information technology and computerization which has become essential for learning. Due to technology advancement online learning is increasingly popular nowadays providing the students with flexibility in terms of gaining knowledge at any time and in any corner of the world. By applying data mining techniques students’ online behaviour can be assessed. To make online teaching better we need to study the pattern of their online usage. In this paper we are applying various data mining techniques to analyse their usage pattern on data collected by students taking online course. Also, a model can be proposed to optimize their learning performance. Data Mining Techniques can be used to optimize the e-learning systems.

1.1 Data mining: Definition and its Types

Data mining is a technique to discover convenient methods from huge databases. As stated in [4] data mining can be defined as “The practice of evaluating earlier concealed, understandable and actionable particulars from large databases and using it to make important business decisions”. Predictive or Descriptive are two methods usually used for data study. To analyse what will happen in future we use Predictive methods whereas if we have to check some similarities, we use Descriptive methods. As data mining is a developing area, the techniques are continuously improving, as new enhanced methods are emerged [12].

1.2 Educational Data Mining

Data mining is getting its hold in many new areas, and education field is one of them. Educational data mining is imminent research done by researchers on data collected from educational settings. Data mining help instructors in not only making their lessons easy but also help in prediction of meritorious students. Various data mining techniques can be applied to find out student behaviour, performance, enrolment etc. Traditional education system can be upgraded by data mining techniques to web-based education system.

Nowadays, Education has become more career oriented and competitive. In this competitive world students want access to knowledge any hour of the day. So, E-learning is upcoming trend. It includes online lectures and notes, online discussions, online tests and many more. Students get enrolled online and pursue their courses. Therefore, we need to optimize their learning techniques for better performance.

2. Related Work

Online learning has been widely adopted throughout the globe. Many prior educational technology innovations have witnessed the unusual route of enthusiastic recognition followed by the widespread dismissal. Will this be the future of online learning or there are chances that this would provide a nurturing ground for latter upcoming innovations in the field of education is yet to be seen. There is a need of learning from the early adopters of this technology to make it a success in later adopters and avoiding the mistakes of the pioneers [22].

The success of the online learning depends on the various factors. Few of them namely: online presence, motivation and participation, actual and perceived learning are seemingly dependent upon course length, target audience, discipline area etc. There is a need of study to identify the factors that contribute to the success and failure of the online course. These factors can help in designing of new courses so that to generate maximum desired output [21]. The advantages of online learning over the traditional learning methods where highlighted in the study conducted for comparison of knowledge creation in the students. It was found that student engagement in traditional teacher centred model of teaching is limited to independent working or working in a small group on a task designed by the teacher. The level of knowledge creation and understanding is better in flipped classrooms where the students are generally studying on their own using available resources on the internet and teacher act as the “Guide by my side”. There is a need to identifying the influencing factors that helps in designing better online contents that can further improve online learning [1, 2].
There are various dimensions like learners, instructors, courses, technology, design and environment have major role in the success of an online learning system implementation. However, all of them have a crucial role in deciding the success of a course but majorly the learner and course dimensions have seen to be more influencing features than others [26].

Student participation in online learning can be analysed by recording their log files. Online test outputs of students were used as primary measure to analyse student performance. Decision Trees technique was applied to analyse the data and it was observed that online learning environment has significant influence upon students’ learning achievements. The study revealed that students who are silent learners perform better in online tests in comparison to other students [25].

Process mining techniques can also be applied for understanding learning processes based on student’s activities trails from MOOC (Massive Open Online Course) platforms logs. Dotted charts, process discovery and conformance checking techniques were found useful and concluded that successful students follow the videos and submit the quiz in structured way then unsuccessful students. Hence, advocated the need of organized course contents and overall structure of the online courses [15].

Though we have today assured learning management systems but there is need of more shielded secure. If online data is collected in real time, it is much expensive. Researchers suggested a new secure online learning system in real time with less operation time which had shielded model for e-learning [14].

There are various methods for analysing the student activity in the online virtual learning environment. General Unary Hypotheses Automaton (GUHA) and Markov chain-based analysis were used to check the online virtual environment students’ activities. They examined in detail the students’ behaviour during their studies [21]. Markov chain based proves to be better as it showed graphical outputs which were easy to understand. Also, it helped to take necessary actions needed to be taken [14]. Researchers checked dropout rate and forecasted student failure from virtual learning environment. They took Statistical data from students and analyzed using machine learning algorithms like K-Nearest Neighbours (k-NN) and applied on Demographic data, Classification and Regression Tree. To get more revised results they used different models and with the help of these models they identified students that were on the verge of failure [8].

Researchers enquired about the relation between student behaviour and its effect on studies i.e. if student gets bore studying a particular subject like mathematics, it affects the result of that subject. They concluded that lack of concern is related to output of the student log records of 412 students to check their behaviour and response towards e-learning [16]. They used k-means clustering to do their research. They found that students were not inclined to do self study [19].

Another study in 2014 used Moodle platform to study how online learning is better than traditional teaching methods. For this, they used various data mining techniques, text mining, statistical and cluster analysis. Their study concluded that learning analytics are influential to improve online learning [9]. They examined MOOC data to improve the design, delivery and assessment. They deduced that software agents can help to some extent to optimize study material, course transmission, decrease drop-out rates [5].

Several algorithms were suggested by researchers to optimize learning data given by educators. They proposed a new algorithm to categorize learning material in a better way, which will indirectly help students to understand their subject better [7]. Wolff used the algorithm GUHA along with online data to examine that which students fail in a given module or which students are on the risk of failure. This will further help teachers to interrupt and guide students who have tendency to fail in final [27, 28]. Romero scrutinized that if final year students wants to improve their grades and achievement then their participation in online discussions could be analyzed using various data mining approaches. They used classification and clustering methods instead of traditional classification algorithms to get results [23, 24]. Some analysts applied text and data mining techniques on automatically recorded online questions and chat messages in live video streaming. Then different data sets were used to study in detail, correlation accordingly. The results showed different behaviour of students when discussions were within them and when discussions were with their teachers. Also, they concluded various worth defining patterns which help to optimize courses [10].
Indian analysts also examined the classification task to evaluate student’s performance and the decision tree method was also used. They extracted knowledge that describes students’ performance in end semester examination. It helped in identifying the dropouts and students who need special attention in earlier stage, which allowed the teacher to provide appropriate advising/counselling [3]. Researchers applied classification prototypes for estimating student’s performance and also did clustering to segregate them based on their perception in online learning. This helped the authors to have a better understanding of learning and enhancing course material [11]. Hybrid clustering methods can be utilized to obtain the Web usage mining (WUM). The WUM process contains three steps: initial processing, data mining and result analysis. The Web logs extracted from the real world Web servers used hybrid clustering methods based on Fuzzy C-means. Cloud computing was also embedded with web mining. The outcome of evaluations expressed that using more efficient algorithms for finding similar users lead to recommender system that provides more interesting advises for online [18]. In a similar study, researchers examined performance predictive model in which data was collected from online student usage logs. Log data got speed up due to result of cloud computing usage. To analyse their final performance five Weka algorithms were used [17]. From the results, course managers were able to distinguish that which students need more care and attention [6]. MOClog was used to scrutinize the knowledge gained through online courses. The main aim was to develop a tool for the analysis and representation of data collected from log files. On this basis, a structure was made with concept map, a solution map and use cases. This was implemented and tested and after several testing rounds, MOCLog was ready for use in various academic organizations[13].

3. Research Methodology

3.1. Participants

The study included 237 students of Bachelors of Computer Applications (BCA) at Post Graduate College, affiliated to Panjab University, Chandigarh. The study was conducted by floating an online course on Data Structures to the students and all the course materials were provided online. The platform used was self-build online module on data structures using google classroom and links were provided for video lectures on YouTube. The students were trained to check and download course related material and videos, to check and post online messages, to participate in online discussions.

3.2. Research Process

The research process as shown in Fig. 1. is divided into two parts: Data Collection and Questionnaire.

![Research Process Diagram]

Fig.1. Research Process

The identified research questions that will be answered by study are as follows: -

- How we can enhance online learning as it is an upcoming trend in present education system?
- What are the factors influencing online learning systems?
- Which factors have greater impact on student performance doing online courses?
To study the factors which influence online learning, the data generated in form of log files can be used. Application of data mining techniques on log files requires storage, pre-processing and analysis of log files which is a tedious task. For our study we performed an in-depth analysis of the log files generated for different students manually and extracted the factors from the log files. After the analysis, a questionnaire was designed that contains information that can be concluded from log files e.g. number of hours online per day, number of questions attempted, time taken to attempt the question etc. The questionnaire consists of 50 questions out of which 15 were related demographics and subject matter and 35 factors were identified from the log files. The questionnaire was validated using Martuz content validity index method achieving average score of 4.2 out of 5. The questionnaire was filled by the students attending the online learning module and data was pre-processed for further use, as we do feature analysis; as Classification and Clustering techniques were applied on this data. Further, we were able to find influential factors by applying decision tree.

3.3. Data Analysis

Before data analysis, we need to do data processing using Clustering. K-means Clustering is used basically to group data into similar elements. So, that we can group some variables on which further predictive model is applied. With this model we will be able to find out factors which have impact on their final grades. Data preprocessing is done, thereby removing all missing values and noise. The parameters found suitable were as follows:

- No. of times student login
- No. of times participated in online discussions
- No. of times messages posted
- No. of times messages read
- No. of times checked given online material
- Total Grade

4. Applying Clustering Technique

The questionnaire was specifically designed based on identified parameters and data was collected from questionnaire designed for the survey. The survey contained 50 questions related to demographic information, academic information and usage behavior during online learning. Total number of respondents to the survey was 157 out of which 98 were complete and found suitable for analysis. The data collected from the survey was pre-processed and analysed using Rapid Miner. Supervised learning Techniques were used for analysis and the results were compared. The results obtained after comparison were used to find the most influential parameters that influence Learning usage of online students to enhance the performance of the students. When clustering techniques were used on this data to classify students based on various factors listed. As shown in Table 1 students were divided into parts called Clusters. First Cluster had 30 students, next 7 and last Cluster had 61.

<table>
<thead>
<tr>
<th>Cluster1</th>
<th>Cluster2</th>
<th>Cluster3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=30</td>
<td>N=7</td>
<td>N=61</td>
<td>N=98</td>
</tr>
<tr>
<td>Total Student login</td>
<td>51.84</td>
<td>50.87</td>
<td>24.66</td>
</tr>
<tr>
<td>No. of times participated in online discussions</td>
<td>4.98</td>
<td>3.00</td>
<td>2.64</td>
</tr>
<tr>
<td>No. of times messages posted</td>
<td>58.05</td>
<td>41.72</td>
<td>17.36</td>
</tr>
<tr>
<td>No. of times checked messages</td>
<td>92.87</td>
<td>25.87</td>
<td>25.12</td>
</tr>
<tr>
<td>No. of times checked given online materials</td>
<td>54.49</td>
<td>52.87</td>
<td>25.07</td>
</tr>
<tr>
<td>Total Grade</td>
<td>84.14</td>
<td>88.14</td>
<td>64.33</td>
</tr>
</tbody>
</table>

Table 1. Means of Clustering
Table 2. Means for Re-clustering Results of Poor-Performing Students

<table>
<thead>
<tr>
<th></th>
<th>Cluster 3-1 (N=43)</th>
<th>Cluster 3-2 (N=18)</th>
<th>Cluster 3-3 (N=61)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of times Student login</td>
<td>32.61</td>
<td>5.24</td>
<td>24.66</td>
</tr>
<tr>
<td>No. of times participated in online discussions</td>
<td>3.00</td>
<td>0.73</td>
<td>2.64</td>
</tr>
<tr>
<td>No. of times messages posted</td>
<td>21.32</td>
<td>6.3</td>
<td>17.36</td>
</tr>
<tr>
<td>No. of times checked messages</td>
<td>32.06</td>
<td>8.17</td>
<td>25.12</td>
</tr>
<tr>
<td>No. of times checked given online materials</td>
<td>33.46</td>
<td>4.57</td>
<td>25.07</td>
</tr>
<tr>
<td>Total</td>
<td>72.19</td>
<td>45.12</td>
<td>64.33</td>
</tr>
</tbody>
</table>

Students in Cluster 1 and 2 performed better, if we calculate total grade. But students in Cluster 2 gave more good results than 30 students of Cluster one. They checked and posted fewer messages than Cluster 1 students but still gave good and better results. Cluster 3 students performed very poorly, a re-clustering was done. Again, dividing them within into sub-Clusters as shown in Table 2. Cluster with N=43 students gave better results and cluster with N=18 did not do much online activities.

5. Decision Tree Analysis for Prediction

By implementing decision tree, we have analysed the predictive model. As we can see that number of times students checked given online material plays important role in student final grade. If it is more than 19.5 then final grade are 78.93 and if it is more than 45.6 then final grade goes up to 89.62. Also, when students checked messages more than 65.6, their grade improve to 87.01 and when they posted messages more than 89 times their grades increase from 71.64 to 89.64. Thus, we can say in that, all these factors (checking given online material, checking online messages, posting messages) are playing important role in increasing performance of students studying online course.

6. Results and Discussion

The discussion here laid emphasis on two views: one is hands on learning online course and other is potential of educational data mining. The study done was in detail about student participation in online activities.
Students who were not very active also gave good results. That means those who just checked and read given material scored better grades. The course floated was made with utmost care according to the student needs. Proper guidelines were also given regarding participation in online discussions but still they preferred to remain dormant and just read given material. Though decision tree analysis showed participation in online activities like reading messages and posted played influential role in prediction, rather than just reading course material only. Using data mining techniques, a teacher can observe and analyse student activities and can detect students who are at-risk. With Decision tree and Clustering together, an educator can enhance performance of poor performing students by timely making them active in online discussions. From the survey it was also observed that students are active on working days and remain dormant in festive seasons and holidays. Thus, teacher needs to make them participate in online discussions during their college working days only. Also, the assignments should be completed within this time span only. This will gradually increase their performance in online learning course. Therefore, data mining techniques can be used to detect at-risk students, and can help educators to learn booming and winning learning styles.

7. Conclusion

As we know that data mining is powerful and robust technique to find out solutions to many areas and is playing great role in learning behaviour patterns of students. From the study done, we can conclude that observing students’ participation in online course help us to depict various data results. Data mining techniques have supremacy to meet the challenges faced by educators who are in constant struggle to improve online learning. If proper measures are taken to improve the diagnosed factors, then surely students studying online courses will come out with good grades. However, predictive model needs to be validated by applying it to other online learning environments, if we want to enhance the performance of students in online courses.

References


Identification of Influencing Factors for Enhancing Online Learning Usage Model: Evidence from an Indian University

Student Online Behaviour in a Virtual Learning Environment”, In Machine Learning and Learning Analytics workshop at The 4th International Conference on Learning Analytics and Knowledge (LAK14), Indianapolis, Indiana, USA.


Identification of Influencing Factors for Enhancing Online Learning Usage Model: Evidence from an Indian University


Authors’ Profiles

Dr Sachin Ahuja holds a PhD in Data mining. His primary research interests are in the field of educational data mining. Specifically, under data mining he is interested in predictions, designing of survey questionnaires, measuring and comparing the academic performance of students and comparison of traditional teaching with flipped and blended learning models. Apart from data mining, his teaching interests include big data, relation database and procedural languages. In addition to this, he is heading the Office of Patent Facilitation where he has facilitated inventors from Chitkara University in filing patents by guiding them in licensing & consultancy. Sachin in his free time practices yoga and explores the city for good vegetarian cuisine.

Prof. S. N. Panda, Director Research, Chitkara University is Ph.D. in Computer Science with more than 25 years teaching and research experience and working towards development of innovative technologies and product based on Internet of things. He has expertise in Cyber security, Networking, Advanced Operating System, he has published numerous papers based on these areas and produced 5 Ph.Ds and currently he is guiding 7 Ph.D. Scholars and 3 ME Scholars. He has filed 5 patents in the relevant area and involved in Internet of things healthcare devices like Portable Intensive Care Unit, Digital Laryngoscope etc. He has developed the prototype of Smart Portable Intensive Care Unit through which doctor can provide the immediate virtual medical assistance to emergency cases in ambulance and won prestigious Millennium Alliance Award from FICCI in 2017 and seed funding for his project. He is also working on a project “Cyber Technology Communication for Women Safety” which is funded by Ministry of Science and Technology, Govt. of India. He has been bestowed with “Adroit Researcher Award” for the outstanding contribution in the field of Education and Research, during the “International Conference on “Interdisciplinary Research for Sustainable Development (IRSD-2016)” in NITTTR Chandigarh and COSMIC Outstanding Researcher Award from COSMIC Journal, Thailand. He is also proud recipient of Teacher’s Excellence Award from Chitkara University, Punjab INDIA. He received Certificate of Appreciation as Productive Member of International Reviewer Board, Informing Science, Vilnius, Lithuania (Europe) 2014 and Appreciation Award by ISI Florida, USA for excellent work as Member of International Board of Reviewer 2015. He represented Chitkara University in TiECON 2017, in Silicon Valley, California, USA.
Puninder Kaur is pursuing her PhD from Chitkara University, Punjab. Her research interests are data mining applications in education. She has been teaching professional courses in computer science to undergrad students for past 5 years.

How to cite this paper: Sachin Ahuja, Puninder Kaur, S N Panda,"Identification of Influencing Factors for Enhancing Online Learning Usage Model: Evidence from an Indian University", International Journal of Education and Management Engineering(IJEME), Vol.9, No.2, pp.15-24, 2019.DOI: 10.5815/ijeme.2019.02.02