Measurement of E-Learners’ Level of Interest in Online Course Using Support Vector Machine

Awais Ahmed Shujrah¹, Samina Rajper¹ and Awais Khan Jumani²*

¹Department of Computer Science, Shah Abdul Latif University, Khairpur Mirs, Sindh, Pakistan; shujrah60@gmail.com, samina.rajper@gmail.com
²Faculty of Science and Technology, ILMA University, Karachi, Sindh, Pakistan; awaisjumani@yahoo.com

Abstract

Motivations: E-Learning is more popular in today’s era; most of the institutes have E-learning systems and also provide distance learning education. On the other hand, some of the main reasons for student's failures in academics are lack of communication with teachers. Problem: This type of gap between student and teachers make big flaws in academics. Due to this type of gap, the student cannot take an interest in any academic course. This study aims to measure the level of student’s interest in any online course using Support Vector Machine (SVM). Objectives: The main objective of this research is to measure the student’s interest level in online courses and classify the e-learners on the basis of their level of interest in the course. Methods: The information from the spring semester of 2019 from the Department of Computer Science and encompass 597 students and 39 diverse courses. SVM technique has been used to filter the data and collected sequence data especially data processing, clustering, classification, regression, visualization, and feature selection. Findings: In the last, our system can detect the quantity of pages per session. In this system, we accomplish the different levels of students with their learning styles. The teacher can also be able to improve themselves with this system. The teacher may change the teaching methodology with their students. Application: It has been created a platform for classifying the students with weblogs interest in a particular course. After classifying the e-learners on the basis of their interest in a particular course will help the teacher/ E-teacher to change or improve the instructional strategy to develop/improve the level of interest of students.

Keywords: Machine Learning, Support Vector Machine, E-Learning, Web Logs, Clusters

1. Introduction

Recently, online learning activity shows a much better result in educational level and during the learning process, it helps a lot. If some student cannot reach school destination, then he/she can easily take advantage of the online learning education. When some of the experienced teachers want to teach each and every student then they can give facility to students through online learning education. In this case, the data mining technique paying an important role in the educational side, it is also known as Knowledge Discovery Database (KDD) Where all students can easily approach their perspective knowledge and their education.¹ As the grooming field of data mining, it shows many better results some years ago and it is known as emerging issues of the computer science field.

Moreover, machine learning (ML) is one of the strong techniques to work automatically. In the educational sector. ML is playing an important role in the system. Obviously, in the recent era education system depend upon technology each and every technology giving an important part. Such as online lecturers, online quizzes, online tests, and online subject material. ML makes it easier to select a subject and take decisions.² On the other hand, it provides more facilities in the educational system. Many of the clustering algorithms have used in ML for
a different purpose. Algorithms have their own work efficiency and results, so the SVM has used for clustering of the objects and text and provides better results with accuracy. During the ML it helps to cluster a lot of records of students and teachers.

This study shows the student's interest during the curriculum study. After that, it helps the teacher can be able to work on that particular student and to give them perfect and proper counseling. A lot of work shows inside the LMS but in this system, it shows the level of student and helps a lot of students during the bachelor level of study. The main and perfect reason for this study is to provide a better quality of education. Because education is our basic need and every student should know their expertise inside any field.

The core benefits of this study are:

- Data mining technique can be used in it for student and course evaluation.
- The main purpose of this study is to use a data mining algorithm for new indexes and metrics.
- These techniques can be easily adjustable with any LMS.
- Such a user-friendly environment the visualize results and communicate with given data.

The organization of this study depends following aspects: section 1 introduction shows the basic background and the machine learning things, section 2 shows the keywords and core text of machine learning and educational things, section 3 shows the previous work of e-learning, section 4 shows the brief description of the work and steps. In the last section 5 shows the discussion and conclusion of the complete study.

### 2. Educational Research Model

Several advantages concerned with EDM via more educational research models, like practical works. At the beginning of the educational data source where all data sets are available and easily generated which makes EDM extremely feasible. In this regard, it shows all records of regular or non-regular students. Such important files of educational sets include in reliable learning tasks, in this regard student performance can be increased rapidly and maybe touch the high peak of educational level. For measuring feasibility with environmental challenges on students' performance so the researcher can easily handle another paradigm of educational research.

The rapid development of the educational system increasing day by day and sharing information on student performance can be affected with better results. In educational research, paradigm shows that the increasing rate of students learning the environment in the real world. A lot of the students take help from educational apps and software. Using those software students can take help and knowledge but some of the things will be still missing that students have not any ability to take their interest subject regarding their course. After that student can give feedback regarding their performance and academic records.

### 3. Machine Learning

Application in AI is known as Machine learning which makes the system capable of to automatically grab, learn and improve with the experience with confusing the programming done for it. Machine learning mission is based on the development of programming in a computer which can access and later on utilize it for themselves for learning. Machine learning starts with data or observation like experiencing directly, giving instruction, so you look in patterns in data and it helps you to make the best judgment in the future depending on the giving example we have mentioned. The mission is to give access to the computer learning automatically without any interference or disturbance for human and then act on it accordingly.

#### 3.1. Machine Learning Methods

We categorized algorithms of ML as an unsupervised or supervised category.

- You can Supervise learning of machine algorithms by applying what has been taken while learning the past with new data being labeled taking an example so you can predict future events. You give it a start from analyzing the trained data and learn algorithm so it can produce an inferred function that can predict the values of output. Later on, the system can easily provide certain targets for the new input since enough training done on it. While you can learn algorithm which compares the result with intended, finding errors and correct so it can enhance and correct the module as needed to be.
- You can use learning algorithms that are unsupervised when the trained information is not labeled or
classified. You use unsupervised learning to know the infer function so explain a hidden data from the data which is not labeled. You can only explore data as the system cannot detect any output and you can from datasets you can draw interferences and then describe and hidden data which is unlabeled.

- Something which falls between supervised & unsupervised learning is known as Semi-Supervised. This can use both the category unlabeled and labeled in training, mostly a minimum amount of data is labeled and the maximum amount of data is unlabeled. We use this semi supervision when we need to find out relevant and skilled sources so we can train them and take knowledge from it. You do not need an additional source to get unlabeled data.

- Action produces by interaction or discovering rewards or finding errors is done by reinforcing algorithms of ML. Search & delay, error and trail can be described as reinforcing learning of machine. A method that allows software providers and machines to get ideal behavior automatically with a limited amount of data so you can get maximum performance. Feedback given to the provider can help to learn a lot and know which action was best. ML analyzing the huge quantity of data can be done. This delivers it faster with a better result, which can help to locate opportunities that are profitable or dangerous; however, it does require recourse and more time to train properly. To make it better and effective you can combine it with cognitive and AI tech with a large volume of data.

ML allows examining the huge amount of data. Normally, it gives fast and most accurate output to increase the opportunities or risks. It also needs extra time and resources to train. The integration of ML with AI and technologies can make it easier to process the huge amount of data. The main function of ML is to target functions, so it can be work accurately as undefined data and unseen data instances. ML can easily target the function which can be (ho) is sometimes called a model. Figure 1 shows the model of the learning process.

In view of marked training examples, the learning algorithm finds the structures or patterns in the training data set. From there, it creates a model that sums up well from that data. Normally, the learning procedure is explorative. Much of the time, the procedure will be played out numerous times by using various varieties of learning algorithms and configurations.

4. Support Vector Machine

SVM is known as for clustering and an effective algorithm to classify the text and make a decision. In other words, it is related to other classification algorithms such as kernel trick which is used for non-linear input spaces. As well as, it shows the importance in every field of technology and used widely in many applications such as face detection, classification of emails, news articles and web pages. After that, it used mostly in text classification and handwriting recognition.

SVM is a delighted algorithm and adequately simple concepts. The classifier discrete data points using a hyperplane with an enormous amount of margin. SVM is known as a discriminative classifier due to this reason. It finds the finest hyperplane, which helps in the new data point's classification. SVM could be handle smartly multiple continuous and categorical variables. To distinct different classes, it constructs a hyperplane in multidimensional space. It produces a huge hyperplane in an iterative mode that used to slash errors. The principle of SVM is to perceive a maximum marginal hyperplane (MMH) that superlative half databases into classes. In this case, the Figure 2 shows the classification of data.

5. Literature Survey

Some related work shows that the rapid development of student learning in their educational environment. In has discussed the recognition system. In this system, the author has worked on a tool which recognizes the text
and makes the proper arrangement of it. The author has introduced the algorithm that recognizes the text in the system. Similarly, has developed a system that categorizes the text with the help of SVM, in this system the author has worked on SVM for collecting better results. After that, the author has collected 67% result accuracy and categorizes five tags. Furthermore, has implemented a system during the students learning process. In this system, the author has implemented a web application and used that application in different schools. After that, the author collected the result which shows the student interest during the study. The author compared that Game-based learning is better than traditional learning methodology. Moreover, has used different and special data mining techniques for the recognition of learning styles on LMS. The author has collected experimental data. On the basis of their results, it shows the better improve technique of LMS. In this system, their technique shows the student interest according to teach them. In has studied the data mining technique in E-Learning tools. The author has collected their results from teaching experiences belongs to the information system course. Some of the data mining techniques were taught inside the course which is part of that course. After that, the author has introduced the learning tool to analyse the data mining techniques. Some studied parts belong a case study with customer switching prediction. Similarly, has discussed the E-Learning objectives, methodologies and limits of E-Learning techniques. The author has focused on a different kinds of tools and techniques in methodology. Some of the issues have been discussed in particular learning, design, and other communication issues. Finally, the author has concluded and design a tool that allows the user to interact with it from anywhere and also mentioned the feedback of users with improved suggestions. Likewise, has proposed a system which belongs to an online educational system. The author focused on campus wise data which can be collected from different sources of campuses. In this system, the author has collected student data from different campuses using the data mining technique. Correspondingly, has introduced the framework of smart cities where all data can be collected online. In this system, the author has worked on smart cities and analyzing big data. A huge amount of data can have handled with smart cities’ approaches. Additionally, has studied the improved mode of E-Learning approaches and that system had efficiently worked on the educational environment. The author worked on the e-learning model and provides feedback from the audience. Moreover, has researched the common challenges which are included within their business environment are sophisticated and unstable. Therefore, people are interconnected with the same organization on the same network. The author used deep learning techniques inside his and collect their results. Furthermore, has developed a unique system where all data can be collected and stored online. After that student can easily choose their desired data with a unique identification. Mostly all students can easily ask any questions regarding their subjects.

6. Methodology

In this case, the main heading of the research is depending upon two main steps. After that, the proposed methodology works on the learning phase and the prediction phase. Where the learning phase is used for student data and their searching interest using web application. Because many of the students have different approaches and different interest levels. Interest level data can be collected from the student selection course. On the other hand, the second phase can be predicting the measurement of an interest course level. These data can be extracted from the local Web server of an LMS and instructors can easily evaluate the student’s interest level.
6.1. Logging the Data

Such types of steps can show the student logging data on the platform of E-Learning. Inside the e-learning module, some fields have been recorded using a web server of E-Learning. It can record some fields of the E-Learning platform. Specifically, python used the special field for data saving inside the record which is Named Tuple. Inside the python Named Tuple can easily work and record the data of the student’s course. Such of the python tornado web servers can be used for the configurations of Log files which are given below.

- Extend_Log_Format
- Cookie_Log_Format
- Forensic_Log_Format

In this case, much of the previous work has been shown inside the application and such requests can be stored on servers. After that, the admin can easily fetch the data from the server. Moreover, for the fetching of record two queries are working inside the application, on is Unique_ID and another one is Sign (+), which easily indicates the first and second record of the user on each time of the request. Furthermore, inside the web application, some APIs have been used which are a direct concern with the webserver of an E-Learning application. During the collection of data, it is an independent platform of an E-Learning application.

6.2. Data Pre-processing

Information in the logbook has a noise like lost values, outliers, etc. The standard will be pre-processed so that the data mining analysis standard prepared. Definitely, the phase extracts stored information received from the first phase. Upper lever finding used and values at risk are taken away. This phase is not done by the eLearning platform and therefore merged with a variety of learning management systems. Logbook created during the last phase strained, hence contain the field below:

- Course ID, Course identification
- Session ID, Session Identification;

Even if, the fields above stated have the facts for the whole eLearning procedure, further catalogues and metrics Table 1 offers in order to effectively assist the assessment of course practice. Hence a lot of metrics are available on the web to analyze the E-Commerce and parallel matrices are missing for eLearning, which have been merged with the Model Learning Management system so the quality of learning management system is assessed and as well as procedure of learner’s coordination.

### Table 1. Shows the class activity

<table>
<thead>
<tr>
<th>S. no</th>
<th>Classification of activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Theory</td>
</tr>
<tr>
<td>2</td>
<td>Assignments</td>
</tr>
<tr>
<td>3</td>
<td>Related material</td>
</tr>
<tr>
<td>4</td>
<td>Video lectures</td>
</tr>
<tr>
<td>5</td>
<td>Books</td>
</tr>
</tbody>
</table>

6.3. Dataset Description

The results of various institutional changes and modifications that custom the Open E-Class and eLearning policy of the GUNET platform is an integrated Electronic Learning Management System. It follows the philosophy of open-source software and supports the Asynchronous eLearning service without restrictions and commitments. Access to the service is through the use of a simple web browser without the need for specialized technical knowledge. The information from the spring semester of 2019 from the Department of Computer Science and encompass 597 students and 39 diverse courses. The information is in American Standard Code for Information Interchange system (ASCII) (formerly intended for use with teletypes, and so the explanations are slightly cloudy and their use is recurrently not as proposed) and is acquired from the Python Required Fields.

6.4. Preprocessing

Logbook formed for prior step is saved, thus that one holds only the fields:

- Course ID
- Session ID
- Uniform Resource Locator (URL)

As defined in the previous Table 1 student are categorized into their categories. On the basis of the categorization of a student, it can have clustered the student’s interest using their activity. After that, we have collected 687 student’s activity data. In this case, we have just shown 20 different course data from different semesters. During the clustering, it focused on assignment submission, quizzes, class discussion, Mid Term, and Final Term.
Furthermore, Table 2 shows the different interest levels of students. During the different semester lot of the students have taken different interests in the subjects. In Table 2 we have just shown the five subjects with different interest levels in a different activity.

<table>
<thead>
<tr>
<th>Course</th>
<th>Assign sub</th>
<th>Quizzes</th>
<th>Class discu</th>
<th>MID term</th>
<th>Final term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to information technology</td>
<td>27%</td>
<td>13%</td>
<td>17%</td>
<td>12%</td>
<td>31%</td>
</tr>
<tr>
<td>Discrete math</td>
<td>21%</td>
<td>19%</td>
<td>7%</td>
<td>5%</td>
<td>48%</td>
</tr>
<tr>
<td>Fundamental programming</td>
<td>45%</td>
<td>43%</td>
<td>55%</td>
<td>67%</td>
<td>89%</td>
</tr>
<tr>
<td>Basic electronics</td>
<td>47%</td>
<td>49%</td>
<td>53%</td>
<td>43%</td>
<td>67%</td>
</tr>
<tr>
<td>Pakistan studies</td>
<td>12%</td>
<td>13%</td>
<td>9%</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>Islamiyah</td>
<td>17%</td>
<td>21%</td>
<td>13%</td>
<td>31%</td>
<td>44%</td>
</tr>
</tbody>
</table>

6.5. Clustering Results

The following pace discussing the recommended approach implicates the assembling of the information. In this period, we executed 2 distinctive assemblings. The initial one collection are courses centered on the suggested metrics termed at the prior pace. The next collection is the undergraduates affording to the courses they have incorporated. Therefore, the suggested approach be responsible for the mentor with perception not merely into the courses, on the other hand also for learners.

6.6. Course Clustering

In this section, we have clustered the course with a different class activity. During the course clustering data mining technique has been used and it works with python numerous student data can be taken for collecting better results. The SVM technique has been used to filter the data and collected sequence data especially data processing, clustering, classification, regression, visualization, and feature selection. The properties of SVM are based on Euclid’s method consists in supposing a small set of naturally appealing axioms, and gathering many other propositions from these objectivities through two collections, subsequently, our aim was to isolate the 39 courses into extraordinary commotion and short commotion. Outcomes found that 9 (29%) of the courses had extraordinary action and 30 (71%) of the short action.

6.7. Student Clustering

In this section, we have clustered the students with their class enrollment and course ID. If a student visits the Web Application, then he should first signup their ID. After that, he visits their subject with different classified activities. In our system, their respected time can be calculated. Because we can have calculated their interest level with their selection. Figure 3 shows the traffic of students during accessing the web application and the SVM can filter the students with their selection and interest subject. Furthermore, it shows the timing when the student accesses their account and how much time he/she can appear on that particular subject.

There is an insignificant cluster of 4 learners consistent to course “Data Structure and Algorithm” (Course ID = 222) which is actuality educated during the previous semester and merely a precise low number of learners ordinarily select and join it. The learner of this course is well-known in the graph (Figure 3) even previously relating the SVM Algorithm, subsequently, the common courses are not being shared within the students. Therefore, the graph illustration of the interactions of the learners previously the collecting can describe quarantined learners and assists the lecturer to determine definite collections of them. The user can also apply the SVM from the Bio Layout surroundings. Subsequently, the assembling is concluded, Bio Layout colors the modules of individually class contrarily in order to visually differentiate each collection (Figure 4).
Moreover, the 3D graph in the form of Bio-Layout. We implement the SVM approach on 687 students and that technique composed from the tornado log file. Students grouped as a pair of fours number can be seen “Professional Structures” are kept back unmoved producing one hard bunch. A sample explanation of this bunch (Cluster 18) is shown in the (Figure 5). We can see the User ID for the respective learner, the Course ID and the quantity of received and disappearance connections for each node (received and leaving links shared operators that visit the same pages), although in this case some of the figures are impacts.

Similarly, Figure 6 shows the detailed results about the cluster 18 because the cluster 18 shows the students record while attending the classes. During the different semesters of students can perform and attend their classes.

In the programming fundamental course, students can take a good interest. After that Linear Algebra course and Calculus Math course, both are similar to the most same exercise.

During the semester student can be able to take interest in a different level. Some students have an interest to solve assignments and course exercises. But some students have no interest in mathematics subject. Due to the lack of interest of student’s teachers can be able to make a meeting with students and talk with them easily about the subject and their interest level. For showing our research on student’s interest we have captured the limited snapshot and focused only on 5 subjects of different subjects.

The additional sample is displayed in Figure 6 where the outcomes for Collection 18 are accessible. The learners of this collection have shared typically the concept (Course ID = 58) of course “Digital analysis of algorithm” and its accompanying lab course (Course ID = 39) that are educated in the third semester. The SVM cluster algorithm brings about to isolate these learns from the whole of 687 learners.

7. Discussion and Conclusion

We display that the recommended metrics be able to compromise an introductory course classification, which in seizure be able to be used as a response for assembling algorithms. These recommend precise activities to mentors so that they can easily update their course outlines and their usability. Moreover, Using the SVM technique the courses can be easily clustered and divided with their results. Many of the students can easily know their interest level than they can pick the right subject and spend their time on it. Furthermore, this system can show the level of students and also suggest them to learn that course within the time.

Explicitly, our system has the succeeding rewards: Customs assembling systems act to detect altered clutches of courses and altered clutches of learners.
Diagram of SVM Cluster assembling algorithm is firm and accessible and pragmatic of foundation period in the field of eLearning. SVM Cluster can be able to load effectively courses into the alike collection centered on the learners’ course visits and be able to discriminate quarantined courses.

The Bio Layout conception tool supports illustrate the assembling outcomes and consent the mentor to advance search the outcomes in a collaborative 3D environment.

We acknowledged comment approximately the system by the mentors. The mentors were knowledgeable almost the indexing outcomes and most of them amplified the excellence and the size of their learning substantial. They upgraded the quality by reordering the learning factual in a constant, categorized and controlled system. They also amplified the extent by entrenching supplementary learning factual. A significant conclusion through the manner of apprising the mentors’ almost our outcomes was that the level of the courses constitutes a significant enthusiasm for the mentors to endeavor and develop their learning substantial. For the reason that of their reciprocal struggle, they respectively need their courses to be exceedingly categorized. Though, rare mentors criticized that the institute of their courses does not support them in taking extraordinary ending marks in the standing gradient. A numerical assessment for metrics of those courses on the initial level and later a mentor's acquaintance to and employment of the metrics would suggest valuable perceptions and support scientifically assess our system.

We similarly plot to supplementary systematize the entire system, which is evolved as a Plug-in tool to systematize that records Pre-processing and collecting paces. The device will run intermittently (every single month) and will be mailed to the mentor's course rank and recommendations. Where that extended contract and the mentors were up-to-date spontaneously by mail nearly the quality of the content of their progressions. In the last, our system can detect the quantity of pages per session. In this system, we accomplish the different levels of students with their learning styles. The teacher can also be able to improve themselves with this system. He may change the teaching methodology with their students. Offline, use of data mining procedures such as pre-processing, conception, assembling, arrangement, worsening and connotation, and determining secreted data outlines. This is helpful for new comer’s students as well as teachers.

References