

# Data Mining Approach towards Students Behavior Assessment Methods for Higher Studies

Sharad Gangele, PhD  
Professor

Department of Computer Science & Application  
RKDF University, Bhopal, India

Kirti Soni

Research Scholar

Department of Computer Science & Application  
RKDF University, Bhopal, India

Sunil Patil, PhD

Associate Professor

Department of Computer Science & Engineering  
RKDF University, Bhopal, India

## ABSTRACT

The quality education to the students is primary point of higher educational institutions. One of the method to achieve high quality education in higher education system is analysis of students behavior with reference to examination performance, mark sheets, abnormal values and other students activities. Data mining methods discovers knowledge from these records for analysis and prediction about students behavior. In this paper, data mining techniques such as association rules and classification are applied to analyze and present a behavior model of students. The students behavior assessment framework is proposed as model for analysis using data mining technique, the model presents the indication to the critical quantities that regulate the students behavior on learning method. The proposed framework can be applied to extract valuable data that shows all characteristic of student behavior by clustering and subdivision of the student behavior large data set.

## General Terms

Behavior Analysis, Data Mining

## Keywords

Behavior Analysis, Data Mining Technique, Prediction Analysis, Association, Classification.

## 1. INTRODUCTION

Data mining concepts in higher education is an innovative emerging research area. The educational data mining (EDM) is related with developing the methods of exploring the unique types of data that come from educational settings [1, 2]. The education has been benefited through EDM from a real E-Revolution as Virtual Learning Environment (VLE) [3]. To share and spread information for both teachers and students, the Internet also assisted e-learning through the available e-resources. The aims of a teacher have changed from teaching concepts into assisting students to learn and how to search, acquire and organize relevant information in different ways [4, 5].

However, learning methods and techniques has changed towards more student-center, problem-specific, challenge-based or cooperative learning. The teacher is instructed to teach or facilitates all the sessions within classroom in traditional learning. In this method, the teacher remarkably talks more than their student. According to the study program and syllabus with the prescribed curriculum, the teachers have to teach the lesson. The teachers take more effort than students, which had become unsuitable to today's curriculum in this traditional learning. The relationship between the teacher and the learner is more impressive and effective through the classifications and categories in Virtual Learning Environment [6, 7, 8]. The Figure 1 represents parameters and model to analyze students behavior. With the rapidly development of the Internet, the

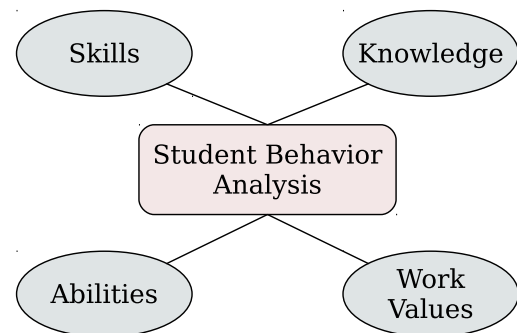


Fig. 1. Students Behavior Analysis Model

web-based educational teaching system has been progressively applied in modern education system, as a significant tool to support learners as well as teachers. In this learning course, there are many profits for information sharing and collaboration between students and teachers. For enhancing student knowledge, learners can take a web-based class at any place, any time and teachers can easily produce their online classes and monitor performance of the student as well [9, 10]. Moodle is a popular Learning Management System

(LMS) which supports educational instructor to create the impressive online courses [11]. However, it do not provide the function to access and estimate motivation behavior of learner. Also log informations of learner in a Moodle can show student's interactions such as reading, writing, taking examination and doing different tasks [12, 13].

With a large collected information daily from e-Learning course, manual data analysis is very difficult. Although there are various tools which assist to report useful information which in turn is very worthful for analyzing pattern behavior of student, the specific features are not offered to teacher need for tracking and evaluating all the activities of students in the class. Data mining techniques is the promising methodology to extract valuable information in this objective [14]. Data mining can analyze related information results and produce several perspectives to understand more about the activities of students so as to customize the course for student learning [15].

## 2. RELATED WORK

Bharadwaj *et al.* [16] proposed the classification task which is applied on student database to predict the students division on the basis of previous database. For data classification, there are many methods and techniques which can be applied to classify dataset, the decision tree approach was used in this research. To predict the performance at the end of the semester examination, informations such as attendance, unit test, seminar and lab assignment marks were collected from the previous databases of the students. This research will assist to the students and the teachers to improve the grade of the student. This research will also work to distinguish those students which required exceptional attention to minimize failure ratios and taking proper action for the next semester examination.

Schaffer *et al.* (2016) [17] analyzed, data from two Massive open online course (MOOC) using multiple methods and multiple visualizations, including local node topography, video watching behavior, global network features and sentiment. They concluded with a concise statement of recommendations:

- For assessment of inclusiveness and forum success, forum networks should be visually monitored over time by the instructors.
- For identifying and assisting orphaned forum students, instructors and course designers should make extra efforts.
- While making predictions, designers of dropout detection systems should take structure of the course into consideration.

A visual and network analytics approach was successful in shedding light on student behavior in this case-study.

Sharma *et al.* (2017) [18] presented a research makes use of the induction algorithm (ID3) decision tree algorithm for prediction of the students final performance based on their previous performances. To make classification, the algorithm uses the criteria of information gain. The more the information gain of an attribute, the more is the split possibility of that attribute. Through this model, they achieved the accuracy of 90%. They hoped that this study will help the students, instructors and the guardians to take necessary action to improve the performance of the students in future. For future work, they like to refine our work by taking more number of example set and come up with more accuracy and other techniques to help students in their educational careers.

Petri *et al.* (2017) [19] proposed that the tweet (or post) given by the agent (people) says about the mood of the person who has tweeted or given his comments. Features can be extracted from various inputs that are coming from various social websites like Facebook,

Twitter, LinkedIn and many more other websites. The same procedure can also be applied for the Sentiment analysis also. The behavior of different people taken in the above said paper is the views of the author itself. Figure 2 represents the process for behavior analysis in social media.

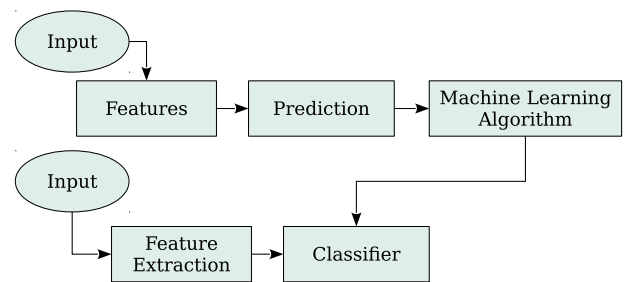


Fig. 2. Process for Behavior Analysis in Social Media

Li *et al.* (2018) [20] presented the natural mathematics of behavior analysis models that generate event records have a general scope regarding the measured aspects of the target behavior. From the simulated event records, they can predict any dependent variable that could have been computed from our data. This aspect allows a freely multivariate analysis of our data. However, they stress that their choice of model, dataset, and model evaluation technique is largely superficial with respect to their aim of demonstrating the use of models that generate event records for a multivariate analysis.

Miller *et al.* (2018) [21] investigated a substantial contribution to the research literature regarding the amount of trait-level variance actually captured by various behavior assessment methods. However additional investigations are greatly needed to replicate and expand upon these findings. The present study on school-based behavioral assessment methods was conducted in order to shed light on the extent to which behavior assessment methods capture intended trait-level variance and determine the degree to which methods utilized impact findings. Support was found for the presence of strong methods effects for all of the assessment methods utilized, which included three of the most common behavior assessment methods used in schools. These findings are consistent with long-standing recommendations calling for the use of multi-source multi-method assessments in the realm of behavioral assessment.

Kularbphetong (2018) [22] presented the prototypes model for the ongoing improvement project to build a model for analyzing learners and learning behavior using educational data mining techniques. The research was divided the study and developed the model into two phases:

- The development of models for analyzing the learning behavior of learners on the e-learning system in the first part was to prepare the log files of the e-learning and e-learning data.
- Student attendance data and transformed the data into a suitable format for further study and development and to develop the model using decision tree and bayesian Networks techniques.

This model can be beneficial to discover the pattern of learning behavior. However, in term of the future experiments, other data mining techniques are be applied to conduct a research to enhance this project and also apply the tool to analyze the learning behavior of learners on the e-learning system.

Shukla (2018) *et al.* [23] presented an outlook for educational excellence indices computation with ranking classification using their proposed index of excellence. Their indices provided the academic

ranking of individual faculty in an institutions of higher teaching-learning and research. The formulae fetch information from big-databases of research and academic contributions for evaluating the rank of a faculty involved in teaching and research in a country. Computer scientist can develop algorithms in order to compute index and ranks in real time scenario. Software developing companies could be motivated for a package development using algorithms of machine learning.

### 3. PROPOSED METHODOLOGY

Data Mining is generalized data analyzing technique from various perspectives to summarize the worthwhile and valuable information results (mined data). Numerous principles are applied to process data mining such as machine learning, deep learning, probability, statistics, neural network and visualization techniques to bring out the knowledge presentation in an easily comprehensible form. In another way, it can also be defined as the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data.

Figure 3 presents the student behavior clustering model which is constructed in this paper. Firstly, the original daily life data of the students are acquired from the database, then, the relevant data is preprocessed and the processed data is counted. After that, the statistical results are analyzed and the relationship between the data sets are obtained. Ultimately, multi-source data sets are integrated. In the data analysis process, this framework can be used to extract valuable data that reflects all aspects of student behavior by clustering and subdivision of the student behavior data set.

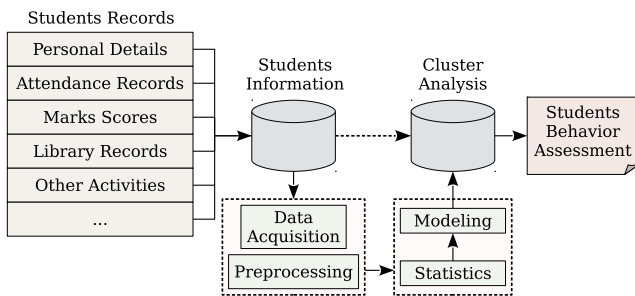


Fig. 3. Students Behavior Assessment Framework

#### 3.1 Data Mining Approach

Association rule mining is applied to discover interesting relationships among variables in databases. According to Agrawal *et al.* [24], an association rule explains a close correlation between items in a database in the form of  $x \Rightarrow y$  where  $x$  and  $y$  are sets of Item set  $I(x \text{ and } y \subset I)$  and  $x \cap y = \emptyset$ .  $I = I_1, I_2, \dots, I_m$  is a Item set of  $m$  distinct attributes. The rule is indicated  $x$  implies  $y$  whereby  $x$  is called antecedent and  $y$  is called consequent. There are two importance thresholds for measurement association rule mining, minimum support and minimum confidence. The support of a rule  $x \Rightarrow y$  is the probability of the Item set  $\{x, y\}$  that means the relevance of the rule and the confidence of a rule  $x \Rightarrow y$  is the conditional probability of  $y$  given  $x$  that indicate the accuracy

of the rule.

$$\text{Confidence}(x \Rightarrow y) = \frac{\text{support}(\{x, y\})}{\text{support}(x)} \quad (1)$$

$$\text{Support}(x \Rightarrow y) = \frac{\text{support}(\{x, y\})}{\text{Total number of transaction in } D} \quad (2)$$

Set of transactions:  $D = d_1, d_2, \dots, d_n$  each  $d_i \subseteq I$ .

Confidence is a significant measure of the association rules to indicate how to strength of the mined rules. If the confidence of the association rule  $x \Rightarrow y$  is 80%, it means that 80% of the transaction the activities of students that contain  $x$  also contain  $y$ , based on users to indicate the specified minimum confidence. Apriori Algorithm is an influential algorithm for association rule mining. The Apriori Algorithm is used level-wise search for frequent item sets, the sets of items that have minimum support. A decision tree is one of the most well known classification algorithms that are commonly used to examine data and induce the tree in order to make predictions. The purpose of the decision tree is to classify the data into distinct groups or branches that generate the strongest separation in the values of the dependent variable.

### 4. CONCLUSION

The proposed model can be beneficial to similar courses to share and discover motivation behavior of students. Data mining techniques can be applied to discover knowledge. Particularly, association rules discovered and the rules can be sorted using lift metric then the rules can be visualized. Then classification rules can be discovered using decision tree. The proposed model as students behavior assessment framework works using data mining technique, it presents the indication to the critical quantities that control the students behavior on learning method. The proposed framework can be applied to extract valuable data that shows all characteristic of student behavior by clustering and subdivision of the student behavior large data set. Finally, using detected outlier analysis all outliers in the data, each one of these knowledge can be used to improve the performance of student. However, in term of the future experiments, a way to generalize the study to more diverse courses to get more accurate results.

### References

- [1] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15 991–16 005, 2017.
- [2] R. S. Baker, "Educational data mining: An advance for intelligent systems in education," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 78–82, May 2014.
- [3] A. do Socorro da Silva, S. R. de Brito, N. L. Vijaykumar, C. A. J. da Rocha, J. C. W. A. Costa, and C. R. L. Frances, "Social network analysis to monitor interactions in virtual learning environment," *IEEE Latin America Transactions*, vol. 13, no. 10, pp. 3482–3487, Oct 2015.
- [4] Y. Wang, "Mechanism of virtual learning environment system," in *e-Learning, e-Education, and Online Training*, S. Liu, M. Glowatz, M. Zappatore, H. Gao, B. Jia, and A. Bucciero, Eds. Cham: Springer International Publishing, 2018, pp. 346–352.

- [5] W. F. McComas, *Virtual Learning Environment*. Rotterdam: SensePublishers, 2014, pp. 110–110. [Online]. Available: [https://doi.org/10.1007/978-94-6209-497-0\\_99](https://doi.org/10.1007/978-94-6209-497-0_99)
- [6] “E-learning: the future,” *ITNOW*, vol. 51, no. 4, p. 28, 2009. [Online]. Available: <http://dx.doi.org/10.1093/itnow/bwp071>
- [7] M. W. Rodrigues, L. E. Zrate, and S. Isotani, “Educational data mining: A review of evaluation process in the e-learning,” *Telematics and Informatics*, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0736585317306639>
- [8] O. Solberg, J. U. Miller, S. Heivang, and M. Bjerke, “E-learning for carers,” *Orphanet Journal of Rare Diseases*, vol. 5, no. 1, p. P13, Oct 2010. [Online]. Available: <https://doi.org/10.1186/1750-1172-5-S1-P13>
- [9] N. Cavus, “Distance learning and learning management systems,” *Procedia - Social and Behavioral Sciences*, vol. 191, pp. 872 – 877, 2015, the Proceedings of 6th World Conference on educational Sciences. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877042815028712>
- [10] R. McDaniel, J. R. Fanfarelli, and R. Lindgren, “Creative content management: Importance, novelty, and affect as design heuristics for learning management systems,” *IEEE Transactions on Professional Communication*, vol. 60, no. 2, pp. 183–200, June 2017.
- [11] R. Calvo, A. Iglesias, and L. Moreno, “Accessibility barriers for users of screen readers in the moodle learning content management system,” *Universal Access in the Information Society*, vol. 13, no. 3, pp. 315–327, Aug 2014. [Online]. Available: <https://doi.org/10.1007/s10209-013-0314-3>
- [12] N. Kerimbayev, J. Kultan, S. Abdykarimova, and A. Akramova, “Lms moodle: Distance international education in cooperation of higher education institutions of different countries,” *Education and Information Technologies*, vol. 22, no. 5, pp. 2125–2139, Sep 2017. [Online]. Available: <https://doi.org/10.1007/s10639-016-9534-5>
- [13] A. Badia, D. Martín, and M. Gómez, “Teachers’ perceptions of the use of moodle activities and their learning impact in secondary education,” *Technology, Knowledge and Learning*, Mar 2018. [Online]. Available: <https://doi.org/10.1007/s10758-018-9354-3>
- [14] A. Ahadi, A. Hellas, and R. Lister, “A contingency table derived method for analyzing course data,” *ACM Trans. Comput. Educ.*, vol. 17, no. 3, pp. 13:1–13:19, Aug. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3123814>
- [15] R. Pellungrini, L. Pappalardo, F. Pratesi, and A. Monreale, “A data mining approach to assess privacy risk in human mobility data,” *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 3, pp. 31:1–31:27, Dec. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3106774>
- [16] B. K. Baradwaj and S. Pal, “Mining educational data to analyze students’ performance,” *CoRR*, vol. abs/1201.3417, 2012. [Online]. Available: <http://arxiv.org/abs/1201.3417>
- [17] J. Schaffer, B. Huynh, J. O’Donovan, T. Hllerer, Y. Xia, and S. Lin, “An analysis of student behavior in two massive open online courses,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug 2016, pp. 380–385.
- [18] G. Sharma and S. K. Vishwakarma, “Analysis and prediction of students academic performance in university courses,” *International Journal of Computer Applications*, vol. 160, no. 4, pp. 40–44, Feb 2017. [Online]. Available: <http://www.ijcaonline.org/archives/volume-160/number4/27065-2017913045>
- [19] A. Perti and M. C. Trivedi, “Article: Markov process for behaviour analysis in social media,” *IJCA Proceedings on National Conference on Next Generation Technologies for e-Business, e-Education and e-Society*, vol. NGTBES 2016, no. 1, pp. 31–33, July 2016, full text available.
- [20] L. Don, H. M. J., and E. Douglas, “The natural mathematics of behavior analysis,” *Journal of the Experimental Analysis of Behavior*, vol. 109, no. 3, pp. 451–474. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jeab.330>
- [21] F. G. Miller, A. H. Johnson, H. Yu, S. M. Chafouleas, D. B. McCoach, T. C. Riley-Tillman, G. A. Fabiano, and M. E. Welsh, “Methods matter: A multi-trait multi-method analysis of student behavior,” *Journal of School Psychology*, vol. 68, pp. 53 – 72, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022440518300207>
- [22] K. Kularbphettong, “Analysis of students’ behavior based on educational data mining,” in *Applied Computational Intelligence and Mathematical Methods*, R. Silhavy, P. Silhavy, and Z. Prokopova, Eds. Cham: Springer International Publishing, 2018, pp. 167–172.
- [23] D. Shukla and A. Alim, “Big data analytics approach using indexing and ranking for excellence in higher education,” *International Journal of Computer Applications*, vol. 180, no. 35, pp. 8–22, Apr 2018. [Online]. Available: <http://www.ijcaonline.org/archives/volume180/number35/29288-2018916878>
- [24] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules in large databases,” in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB ’94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 487–499. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645920.672836>