Analysis of Students'Critical Thinking Skills Using Data Mining Approaches (Survey Based Research)

Dr. M. Pushpalatha (Prof. and HOD) and Raya Bandyopadhyay
Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Kattankulathur Campus, Chennai – 603203, Tamil Nadu, India.

Abstract:- Our aim in this project, will be to identify Key Performance Indexes that can define a student's level of comprehension after studying certain classes and how we can apply those Key Performance Indexes in classification or even use them to calculate the success rate of skills, in universities or even place specific students in areas where they can succeed. We can also analyse which data mining algorithm gives us the highest accuracy based on our data and, address some of the open problems we may encounter as we go along, based on existing research literature.

Understanding the learner's in-depth thinking process after a lesson or series of lessons, will give us more information about where the student is lacking or whether the skills are lacking, in the event that most students seem to lack a certain pattern. This shall enable more fluid methods for students and academics to be classified into a system, we can categorize them based on class performances or regular assignments, and find a system which shall give us an understanding about the grasp of a particular student in a certain subject and eventually, the group of students performing well in certain subjects can be placed in opportunities which shall enhance their skill sets and help them pick a customized career for them.

The use of multi-phase analysis and cluster analysis is intended to be based on data on which Key Performance Indexes will be determined at the end. Based on these determining Key Performance Indexes, we can access important information and, if possible, present it on a working dashboard.

I. INTRODUCTION

In this moment, there is an overtly increased advancement of an online based education system, where any academically involved activity is being followed or tracked and are being maintained in multiple databases, log files, personal profiles, etc. Online based learning is constantly producing large volumes of data describing the ongoing communication between training and teaching, electrical systems and students. In addition, the proliferation of student information and behaviours built into the online based learning system is a growing problem as uncontrolled information can give erroneous ideas and opinions without providing any clear information. Several studies suggest different ways in employing data mining algorithms and techniques in the online based education and learning system.

It is strongly believed and agreed upon in various instances that data mining techniques can be employed in online based learning systems to read and understand academic profiles of students, predict and improve their quality of work, provide training according to academically demanding needs of the students, and most importantly, follow and understand the critical thinking process of the student. This can be made easier by taking regular students' tests and tracking their attendance patterns, which is why relationships can give us insight into what a student's motivations might be and where a student can be placed to achieve better school outcomes.

The performance of students in an institution and various departments is challenging and depends on a number of factors, such as a student's academic performance, communication skills, capabilities, problem solving and disability factors, etc. Work Integrated Learning is a mechanism to enhance performance practice and development of a student in their readiness skills, especially within graduates. The ability to predict student performance using data mining techniques will assist potential employers in placing students in an integrated learning career. The research applies data mining techniques to predict if the student is critically thinking correctly and if they will pass or fail the course. Work Integrated Learning in certain Universities refers to work-based learning activities at an approved industry workstation and the experience that integrate theory and practice. Application of data mining algorithms are helpful to determine information that can be used to establish pedagogical basis for taking educational decisions. The concern that whether a student shall flare well or not in the semester involved or shall prove to be below an average mark, needs to be predicted and factors need to be considered accordingly. Predictions are important for identifying and assisting in student's performance for referral. Classification algorithms using a data mining tool are then applied to the data set for analysis and student performance. The patterns to be discovered in the application of educational data mining techniques will be used to enhance decision making in student placement in any field, further. The prediction of student performance in their final semesters, mainly will help employability practitioners and student for higher studies for proper progress. The research aims to employ data mining algorithms of classification and clustering techniques to interpret potential and useful insights from the raw data given. The predictions of student performance in a university help in identifying outstanding students for industry placement and other higher study allocations. The marks acquired by the student during the

emester decide their future for allocation in the various environments.

II. RELATED WORK

- ***Towards Accurate and Fair Prediction of College Success, Evaluating Different Sources of Student Data:*** In higher education, predictable analysis can provide information that can be applied to a variety of stakeholders such as administrators, educators, and students. Different sets of features are often used for different guessing tasks, e.g., student activity prediction logs and registrar data to predict long-term college success. In terms of fairness, the use of institutional data consistently undermines the number of students who were previously disadvantaged than their peers, while LMS data tends to overestimate some of these groups more often. Combining the two data sources does not completely eliminate bias and still leads to high levels of stigma among disadvantaged groups. This analysis serves to inform the cost-effective and equitable use of student data in predictive analytics applications in higher education.

- ***Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques:*** Technology paired with innovation enable several higher education institutions to employ various forms of academic curriculums - video based teaching and learning is one such technique. Exploratory data analysis of the digital footprints left behind in this online based curriculum enables to completely grasp the efficacy of this type of program. Video-based online interaction and learning and transformational teaching can enable several improvements in a student's academic performance. Knowledge gained from the academic information system, learning management system and mobile applications were reviewed and analyzed with the help of classification algorithms. In addition, data modification and pre-processing techniques were performed to gain more compact features. In addition, genetic research and partial analysis were performed to further reduce traits.

- ***Developing classifier for the prediction of students' performance using data mining classification techniques:*** Data mining is employed in academically involved institutions to predict a student's implementation in curriculum using classification strategies. These techniques are used in student traits to identify logical patterns that can be used as a basis for speculation. Digital student information being highly accessible and an overtly gain in the coursing power of computer systems make the whole process a reality. There is a lot of research done in this way to prevent major student failure. Student features collected from various sources were previously reviewed, which were later introduced to Put to select features and finally read and tested.

- ***Data Mining and Visualization of Large Human Behavior Data Sets:*** Traditional methods of studying human behavior such as surveys and hand-to-hand collections are expensive, time-consuming and therefore cannot be easily used on a large scale. In recent years a growing number of digital social media activities - for example social networking, email and credit card purchases - have given us new resources to study our behavior. In particular, smartphones have emerged as a new tool for collecting data about human activity, due to its sensory capabilities and ubiquitous presence. This thesis explores the question of what we can learn about human behavior in this rich and ubiquitous mobile sensor data.

- ***Data Mining Methods for Detecting the Most Significant Factors Affecting Students' Performance:*** In the last ten years, we have seen a surge in numerous real life, practical studies to understand and explore student data methodologically so as to understand and predict their performance and fulfillment. Student's academically flaring well involves higher education institutions' major interests and enables newer opportunities. Keeping at par with this decision, there is a huge need to oblige the surge and enable methods that study and understand student performance. It contributes to the improvement of the quality of education, and, in turn, encourages educational institutions to gain and gain unparalleled knowledge from the data collected by their students.

- ***On the application of data mining algorithms for predicting student performance, a case study:*** This paper explores the use of data mining methods, namely classification and aggregation, in the data of students graduating with Information Technology degree from a mixed university of study. There is a certain part that employs the decision tree algorithm which is further associated with prediction of a student's performance at the final stage of the code that is involved with social variables and results obtained from the initial circa (i.e. preparation) of the learning program. The classifier is used as well to search for important insights that shall prove as worthy key performance indicators of the academic fulfillment of a student. The conclusive results only bring forward how highly data mining shall be recommended in involving students, academic advisors and university management, to ultimately put these insights into implementation and henceforth, improve the quality of any educational process and eventually each learning experience of each academic.

- ***Prediction of Work Integrated Learning Placement Using Data Mining Algorithms:*** Educational data mining is used to study data and discover new trends or patterns that can be used to gain insights and hence, take certain decisions. The classification algorithms are employed so as to predict student's performance based on academic datasets and their exploratory analysis, and hence, involving an integrated system of learning and placements. The decision tree from the forecast shows students who may be ready to be placed in the workplace. The study compares different data mining techniques to differentiate a student based on the data set for the semester before the final exam.

- ***Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining:*** This paper analyzes the methods that students partake to interact with academically involved instructional video groups, and their consequences on their academic fulfillments. In this study, auto-emerging video studies were

categorized into multiple pedagogical lists. It focuses on explaining the student's approach to watching the instructional video, and how they navigate the video sequence of that video, in order to predict whether the academic can successfully process and fulfill the session conducted through video. Student video clicks are collected and sorted. It employs multiple data mining algorithms using K-nearest Neighbors and Multilayer Perceptron algorithms to predict academic fulfillments of students. This approach may help teachers to understand how students view instructional videos. This shall be implemented in any unethical or illegal methies that the student might be pertaining to while effective education sessions through videos and also enable instructors to take necessary preventive steps in order to avoid such situations.

- *Predicting academic success in higher education, literature review and best practices:* Student's success stories invite unending opportunities for institutions that they are involved with and are often considered as the tipping factor in many a situations for the institution themselves. Early detection students who are not doing so well or at the edge shall enable to write more favorable factors to such metrics and might even lead to prevention of such failures. This paper aims to hand hold faculties and involved personnel to be able to detect such anomalies and predict results of students and hence, be able to favor towards a more successful batch of students if they are willing to. As a compliance, the literature has been revisited, and the forthcoming insights are integrated into a systematically modelled procedure, in which possible decisions and limitations are fully integrated and defined as well as controversies. This procedure shall enable educators to fully understand and unleash their potential on a group of students who can be understood through the various data mining procedures and hence, enable better results.

- *Adaptive Test Design with a Naive Bayes Framework:* Bayesian image models are frequently employed to derive certain student-based models from specifically related data. Quite a number of predefined models are at our hand available to model certain Bayesian image models from pre-existing student or academic datasets. These models can even put certain observations and learnings of an individual to test at a critical level. In fact, such data often contain deficit values and, in some cases, deficits in excess of the values shown. It determines the employability of a simple heuristic choice of test item based on the level of uncertainty. The uncertainty of an object is based on its initial chances of success, too, because of its complexity. The conclusive insights depict that the model's performance is affected and that heuristic performance definitely leads to better results.

III. PROBLEM DEFINITION

It is a challenging task for both the student and their mentor to understand what field they can excel in, especially in the web-based education system. Understanding the learner's in-depth thinking process after a lesson or series of lessons, will give us more information about where the student is lacking or whether the skills are lacking, in the event that most students seem to lack a certain pattern. To make it easier for students to be organized systematically, we can differentiate ourselves based on class performance or general activities, and find a system that will give us insight into a particular student's performance in a particular subject and ultimately, a group of students who excel in certain subjects. The use of multi-phase analysis and cluster analysis is intended to be based on data on which KPIs will be determined at the end. Based on these determining KPIs, we can access important information and, if possible, present it on a working dashboard.

IV. METHODOLOGIES INVOLVED

The rationalizing in order to conduct an organized preview based on relational system is to solely find suitable procedures for the pre-existing parameters, to bridge any shortcomings in the research conducted beforehand and we find new material to enable the formation of concrete insights and conclusions. The purpose for an organized preview in the ongoing literature is to solely uphold any dubious statements that might be put forth during the process. Next, we shall define the methodologies and procedures that help drive our research forward and help us in progression towards conclusions and results. We are hoping this shall also help us define certain boundaries and help us also identify any limitations that might be present within the research.

A. Probable factors to be considered:
- Relevant questions, descriptive and quantitative both, shall be pertinent to explore the current literature work that helps in understanding and predicting academic performance of students.
- The utmost frequently employed parameters are cumulative grade point average and any internal assessments conducted during the progress of semester. The primary rationale as to why most scholars employ cumulative grade point average is due to its notable contribution to any work or higher studies related institution. This parameter shall also be qualified as a calculating factor for academic potential of an institution.
- Next, the utmost frequently employed parameter is population testing and external testing. Academic demographics concentrate on gender, age, family background, and disability. The rationale as to why scholars often employ demographics as gender is due to the reason that male and female academic styles of learning and teaching are often extremely of dissimilar methodologies. Male academics are known to have extremely efficient and optimized methods of grasping the knowledge and having a better understanding of some of the topics, on the other hand, females are known well to complete due diligence in their methodologies and are rarely found to be slagging, instead they are known to burn the midnight oil quite some nights!
- There have been several approaches that have been made by scholars which involve several psychometric parameters of an academic personnel. The psychometric parameters that are employed are any interest that an academic take in a subject, the enthusiasm to grasp knowledge, the amount of time spent by oneself to study, and family support. These parameters have been employed to develop a structured

system that is efficient and optimized and simple to implement. While this helps, in assessing an academic's trends of behavior or even their choices and the possible reasoning behind them; however, they are not employed as usual as we would like since they can only give qualitative insights and in the research world, quantitative insights are preferred.

B. Algorithms involved:

- **Decision Tree**

Decision Tree is one of the extremely frequent methodologies employed to predict academic performance of students. It is largely employed due to its simpleton ways of being able to understand large or smaller datasets and their structured formation, and eventually predict the parameters as required by the user. Decision tree models are effortlessly comprehended and the whole procedure of their formation can be transformed into conditional, if-then rules. The academic performance of the students is mined from the datasets that are defined in the online based education system.

- **Neural Network**

The Neural Network one of the second most frequently employed methods in mining data from academically involved backgrounds. One of the advantages of neural network can be listed as, its proactive ability to detect all and any relationships and equations between the predictive variables and their possibilities. Neural networks are also known to have the ability to recognize completely mapped out relationships between complicated indirect equations between the predictive variables or indices. Henceforth, the neural network process is labelled as one of the better methods to predict required variable values in a dataset.

- **Naive Bayes**

The Naive Bayes algorithm is an eminent option for the scholars to employ in the prediction business. The purpose of the projects employing this methodology is solely to search for the by far most efficient technique for predicting academically involved performance of students in a compare and contrast manner. The literature only depicts that Naïve Bayes involves all and every available variable present in the dataset. Furthermore, each attribute was analyzed exploratively so as to get a clarified understanding of the independence and the quantitative factor of the attributes individually.

- **K-Nearest Neighbor**

K-Nearest Neighbor is one of the widespread algorithms that has been involved widely in a whole lot of projects with excellent accuracy and good amount of fulfillment. The K-Nearest Neighbor algorithm's perspective has definitely taken minimal amount of time to categorize and classify academic performances of the students into various clusters. It has also shown to provide detailed patterns among the population progress of academically involved students in the higher standards.

- **Support Vector Machine**

Support Vector Machine is one of the supervised methodologies employed for classification of datasets. A few projects employ this particular algorithm since its proven to be extremely well-suited for smaller scaled datasets. Generally, the compilation time and execution time is truly minimal and hence, the time complexity is optimal and even the integration structure is highly favoured.

V. FUTURE SCOPE

The inclusion of meta-analysis in forecasting academic progression of students should rationalize and encourage further involvement in the particular field of studies. It shall enhance the online based education organization to monitor student performance in a structured manner. Although the reported findings are based on an existing literature, i.e., data directly linked to students only, any additional available data may be included in the analysis, such as educator data or parental data that may be relevant and may lead to discovery of new features in the study.

VI. CONCLUSION

Prediction of academic performances are beneficial for both learners and faculties involved with the learning in order to make progress in the system and insights that shall help model the system optimally. This project tries to understand all the methodologies that were involved in the educational system analysis. However, some algorithms improve the performance of categories when used with advanced processing functions such as sorting and re-measuring data, but others do not. Little is known about the complete use of different data sources for all speculative activities and the accuracy of your guesses in relation to the diversity of different people. Research data makes a limited contribution to accurate predictions, as, they are often given to study participants themselves, and that is why information is often biased, leading to inaccurate predictions that are clearly inconsistent with real results and therefore, lead to lower accuracy, or false predictions. Maximum scholars used the cumulative grade point average and internal assessments conducted throughout the semesters as base datasets. While in forecasting techniques, the classification and clustering methods are widely used in the area of academically inclined data mining. K-Means, while a widely used algorithm proves unfit for our dataset with only 48 percent accuracy and hence, other algorithms should be explored in order to gain maximum accuracy. Among the classification and prediction techniques, Neural Network and Decision Tree are the most frequently employed methodologies in the mining of knowledge from academically involved situations. Conclusively, among certain red lights there are plenty of paths to worth pursuing to reach content results.

REFERENCES

[1.] D. Jackson, "The contribution of work-integrated learning to undergraduate employability skill outcomes," Asia-Pacific J. Coop. Educ., vol. 14, no. 2, pp. 99–115, 2013.

[2.] Zafra and S. Ventura, "Predicting Student Grades in Learning Management Systems with Multiple Instance Genetic Programming," no. Mil, pp. 307–314, 2009.

[3.] K. Pal, "Classification Model of Prediction for Placement of Students," Int. J. Mod. Educ. Comput. Sci., vol. 11, pp. 49–56, 2013.

[4.] N. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in 37th ASEE/IEEE Frontiers In Education Conference, 2007, pp. 7–12.

[5.] AZIZ, N. ISMAIL, and F. AHMAD, "MINING STUDENTS'ACADEMIC PERFORMANCE.," J. Theor. Appl. Inf. Technol., vol. 53, no. 3, 2013.

[6.] S. A. Kumar, "EFFICIENCY OF DECISION TREES IN PREDICTING STUDENT ' S ACADEMIC PERFORMANCE," Comput. Sci. Inf. Technol., vol. 2, pp. 335–343, 2011.

[7.] S. Milinković and M. Maksimović, "USING DECISION TREE CLASSIFIER FOR ANALYZING STUDENTS ' ACTIVITIES," JITA, vol. 3, no. 2, pp. 87–95, 2013.

[8.] J. Ghasemian, M. Moallem, and Y. Alipour, "Predicting students ' grades using fuzzy non-parametric regression method and ReliefF- based algorithm," Adv. Comput. Sci. an Int. J., vol. 3, no. 2, pp. 43– 51, 2014.

[9.] O. State, "Mining Parent Socio- Economic Factors to Predict Students' Academic Performance in Osun State College of Technology, EsaOke," Int. J. Eng. Res. Technol., vol. 2, no. 12, pp. 1677–1683, 2013.

[10.] M'hammedAbdous, W. He, and C.-J. Yen, "Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade," Educ. Technol. Soc., vol. 15, no. 3, pp. 77–88, 2012.

[11.] Lara, J., Lizcano, D., Martínez, M., Pazos, J., and T. Riera, A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA. Computers & Education, 2014. 72(1): p. 23–36.

[12.] Abdous, M., & Yen, C.J. (2010). A predictive study of learner satisfaction and outcomes in face-toface, satellite broadcast, and live video-streaming learning environments. The Internet and Higher Education, 13(4), 248–257.