A SURVEY - GPA PREDICTION ON ACADEMIC PERFORMANCE OF STUDENTS

[1]VEGOLUM LAHARIKA, [2]Dr.S. VISWANADHA RAJU

[1]M.tech Scholar, [2]Professor , Department Of CSE,

JNTUH UNIVERSITY COLLEGE OF ENGINERING, JAGTIAL, T.S., INDIA

Abstract: Research in education has evolved to include learning analytics. Learning analytics provides valuable insights about education outcomes and collects data on learners. This article aims to review the research literature that has been published on learning analytics prediction models over the past five-years. This work examines the topic of learning analytics. This includes categorizing based data set, prediction algorithms, and factors prioritised for prediction. This paper also covers the applications, life cycle models and vital challenges for improving the performance of the Learning Analytics System. This study may provide details about learner analytic and possible research lines for researchers to create new models.

1 INTRODUCTION:

Educational data has increased in number rapidly. There is an increasing need to analyze this huge amount of data. Educational Data Mining (EDM), was created to address this problem. Educational data mining refers to the use of data mining tools and techniques in order to analyze data within educational institutions. Educational data mining is a new field that has helped the educational sector develop new methods of teaching and learning. The area of educational mining has numerous papers. This area of research has many benefits for education. Educational data mining (EDM), a technique used by educational institutions for deep and extensive knowledge, is used to improve the quality of its evaluations, planning, and decision making in educational programs. EDM is a tool that academic programs use to find hidden patterns in data. These patterns can be used easily to predict student behavior or

performance. This will enable administrators to allocate resources and buildings more efficiently [2].

Universities are using a variety of methods of data mining to analyse educational reports that are stored in the educational institution like student enrollment, performances, teacher evaluations, gender distinctions and many other. Data mining can provide an institution the required information needed to plan a better amount of students' enrollment, the number of students who drop out and early detection of poor students and to effectively assign resources using an exact estimation.

2 RELATED WORK:

Pauziah Mohd Arsad, et. al. This method, called Artificial Neural Network (ANN) is used to determine the level of academic performance for students. The cumulative grade point (CGPA) is utilized as a measure. The information needed to complete the project comes from the electrical department of Teknologi MARA University, Malaysia. The first semester performance of students is considered the predictor variable for input (Independent variable) and the eighth year grade points for the semester are used for the outcome variable (Dependent variable). The research was conducted at two entry levels: Diploma and Matriculation intakes. The models' performance was assessed by using the coefficient of correlation R as well as the Mean Square Error (MSE). The results of the study demonstrated that the important subjects in semesters one and three are able to exert a significant influence on what is termed the ultimate CGPA.

Midhun Mohan et. al. They typically employ two types of methods for the general estimation of students' performance over a vast amount of data. The two techniques used include Learning Analytics and Predictive Analytics. Learning Analytics mainly deals with the collection of data and data preprocessing, where the necessary data needed to build predictions models is gathered from CBSE schools. They employ the MySQL server to store the vast quantity of data. In the pre-processing process, cleansing, data transformation, and so on. is done using the Apache HIVE framework. After the

data is processed, needed information is uncovered by the MapReduce algorithm within the Hadoop framework. Then is the Predictive analytics phase where actual predictions are created using the multi linear regression models.

Madhav S. Vyas et. al. A decision tree model is used to predict the student's academic progress. The information needed to build this model are collected. Then data processing is carried out where continuous values are transformed to discrete numbers and null values are removed. After that, by applying the CART algorithm on the data, the decision tree model for prediction is constructed. The students who have poor performances are eliminated.

Alana M. et. al. In this case, they're focusing on two methods: Regression and clustering. The use of clustering to prepare data for processing and regression is employed to predict. When clustering, similar data's are put together, despite the fact that data classification is being performed using the regression technique. the model is used to make prediction.

3 CRITICAL GAPS AND POSSIBLE REMEDIES

Machine learning and artificial intelligence are being studied in the field of education sciences as well as every area of daily life. But, the research differs with respect to its goals, the methods employed, the datasets used as well as evaluation criteria and the validation strategy. From the perspective in machine learning though each study aids in and improves student performance assessments However, the different approaches make it difficult to implement the results of these studies in real-world situations.

Discussions on various topics were conducted to offer an overall view of students' performance prediction studies in light of the findings that were presented earlier. Datasets can have a major influence on the goals of research and directly impact the subject of research, the selection of ML model and the metrics used to evaluate. The use of questionnaires is not recommended to estimate the duration of a study due to their reliability and the fact that they are discrete training events once they're collected. However, they are still able to

be considered in research about the effects of personal and family preferences on the student's performance.

The data from online sources have been extensively utilized in in-term, end-of-term and drop-out studies. The analysis of students' enthusiasm for the courses at a specific level as well as time-series data in these datasets allows artificial intelligence and machine-learning models to gain more relevant data and generate more effective outcomes. Furthermore, they've reduced the number of participants which the surveys have to be able to reach, as well as the time, effort and the cost of getting data in digital environments that are part of the infrastructure of schools.

Wang et al. [10] The research has gone one step further, using data that includes the observations of students' activities as well as the way they learn on campus. Thus it is possible to determine, in the simplest method, the location and many hours students spend on campus on campus, the impact of books they purchase from libraries, and other factors could be studied and predictions of student performance can be made based on this information.

4 CONSTRAIONS TO BE EVALUATED

The application of regression and classification tasks in studies of student performance prediction is also a function of the nature of the data gathered and the goals of the research. The widely used Student Performance Dataset [15] allows researchers to conduct investigations in the areas of classification and regression using the exam results available. In contrast in different databases [11] [16], results were classified on the basis of categorized performances or grade of the student, and this can make it difficult to the process of conducting the regression analysis using these data sets. The information contained in the data could enable researchers to conduct predictions studies within the domain of classification (i.e. students' decision to withdraw from the class, final success such as pass/fail) and in the regression area (i.e. the end-of-term achievement (exam point).

regardless of whether the research studies are in classification or regression or classification, they will be classified within the five broad areas mentioned above (estimation of drop-outs from courses and end-of-term level of success prediction as well as in-term prediction of success, students' risk-based identification and the future success estimation) Different questions could be addressed from an educational standpoint:

Do predictions of drop-outs contribute to the student's level of success?

What does the expectation of success at the end of the road contribute to the development of students' self-confidence and education?

Are the predictions for performance in the near-term period of time allow enough time for a contribution for the student?

What is the earliest time to make risk predictions in relation to the courses taken by students help them?

Alongside the impact of the data on studies, less errors and more easily interpretable results from the classification research make them more useful in this area. However the analysis of specific results from regression tasks can make it difficult to assess the results as each of the samples is unique in its error. The findings from the systematic review of literature showed that the various reasons mentioned above led to the use of research on classification (62 percent) to be much more than regression research (38 percent).

The datasets that are considered and the problem domains affect ML models used in these research. If the ML models are examined in this manner, it is apparent that the capacity that neural-based systems process and learn from a significant amount of information and to produce results that are satisfactory is a crucial factor to take into study in a majority of research studies on student performance prediction [15][16][17]. In addition it has been discovered the usage of neural network, that learn through recollecting previous experiences when studying data, particularly in time-series datasets such as online data sets, is becoming common place

5 CONCLUSION

As data growth increases it became difficult for individuals to analyse and process huge amounts of data, but thanks to machine learning and data mining methods, the analysis of large amounts of data has become much simpler. These techniques are not just beneficial in the analysis of data, but also aid in predicting future trends with which massive progress can be achieved.

6 REFERENCES

[1] M. Al-Razgan, A. S. Al-Khalifa, and H. S. Al-Khalifa, "Educational data mining: A systematic review of the published literature 2006-2013," in Proc. the 1st International Conference on Advanced Data and Information Engineering, 2013, pp. 711-719.

[2] F. Siraj and M. A. Abdoulha, "Mining enrolment data using predictive and descriptive approaches," Knowledge-Oriented Applications in Data Mining, pp. 53-72, 2007.

[3] Q. A. Al-Radaideh, A. A. Ananbeh, and E. M. Al-Shawakfa, "A classification model for predicting the suitable study track for school students " International Journal of Research and Reviews in Applied Sciences, vol. 8, 2001.

[4] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance " International Journal of Advanced Computer Science and Applications, vol. 2, 2011.

[6] A. Nandeshwar and S. Chaudhari. (2009). Enrollment Prediction Models Using Data Mining. [Online]. Available: http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Proje ct.pdf

[7] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," Cybernetics and Information Technologies, vol. 13, 2013.

[8] D. Garc´ıa-Saiz and M. Zorrilla, "Comparing classification methods for predicting distance students' performance," The Journal of Machine Learning Research, 2011.

[9] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students," presented at the the 1st International Conference on Educational Data Mining, 2008.

[10] Z. J. Kovačić, "Early prediction of student success: Mining students enrolment data," presented at the Informing Science & IT Education Conference, 2010.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten," The weka data mining software: an update," presented at the Special Interest Group on Knowledge Discovery and Data Mining, SIGKDD, 2009.

[12] E. Boretz, "Grade inflation and the myth of student consumerism," College Teaching, vol. 52, 2004. [13] W. H¨am¨al¨ainen and M. Vinni, "Classifiers for educational data mining," Handbook of Educational Data Mining, 2010.

[14] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

[15] Pauziah Mohd Arsad, Norlida Buniyamin and Jamalul-lail Ab Manan. "A Neural Network Students' Performance Prediction Model (NNSPPM)" IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), 26-27 November 2013.

[16] Midhun Mohan M G, Siju K Augustin and Dr. Kumari Roshni V S "A BigData Approach for Classification and Prediction of Student Result Using MapReduce", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015

[17] Madhav S. Vyas and Reshma Gulwani. "Predicting Student's Performance using CART approach in Data Science", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[18] Alana M. de Morais, Joseana M. F. R. Araújo and Evandro B. Costa. "Monitoring Student Performance Using Data Clustering and Predictive Modelling",
978-1-4799-3922-0/14/$31.00   ©2014 IEEE

[19] Huda Al-Shehri, Amani Al-Qarni, Leena Al-Saati, Arwa Batoaq, Haifa Badukhen, Saleh Alrashed, Jamal Alhiyafi and Sunday O. Olatunji. "Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)

[20] David de la Peña, Juan A. Lara, David Lizcano, María and Concepción Burgos, María L. Campanario. "Mining activity grades to model students' performance", ICEMIS2017, Monastir, Tunisia

[21] Febrianti Widyahastuti and Viany Utami Tjhin. "Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron",
978-1-5090-4688-1/17/$31.00 ©2017 IEEE

[22] Xiaofeng Ma and Zhurong Zhou. "Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree",
978-1-5386-4649-6/18/$31.00 ©2018 IEEE.

[23] Sagardeep Roy and Anchal Garg, "Analyzing Performance of Students by Using Data Mining Techniques A Literature Survey", 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)