Student Performance Prediction via Data Mining & Machine Learning

Abhishek Roy[1], Akhil Sharma[2], Anmol Singh[3], Hari Shankar[4], Prof. Sahana MP[5]

[1-5]Department of Computer Science Engineering, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

---***---

Abstract - *Whenever the word 'Education' is brought into the limelight, majority of the focus is given towards the student, it's a good thing that people are trying to help the 'Learning' aspect; but in order to do that, we often forget about the 'Teaching' aspect, helping the teacher(s) in their method(s) of teaching is equally important as method(s) of learning.*

*Key Words***:** Student, Performance, Prediction, Data Mining, Machine Learning, Decision Tree, Accuracy

1. INTRODUCTION

Nowadays we face a lot of work when it comes to the teaching department, a lot of work in the back-end keeps on going about which the student has no knowledge. And this is our effort to ease down some work so that the work-load is a little less in the future. Using a decision tree algorithm [1][9], the teachers will keep an eye on the grades of the children and with the help of the student performance prediction, would be able to keep an eye on the students who need a little extra attention. One of the things that we often overlook in the Education System is the contribution of teachers in helping the next generation finding the best out of themselves.

2. LITERATURE REVIEW

Huseyin Guruler, Ayhan Istanbullu & Mehmet Karahasan [1] have said that Data mining is one example of knowledge discovery, which is used to locate meaningful and valuable patterns in enormous amounts of data. All of the tasks involved in the knowledge discovery process are maintained together with this software system. The research was based on data collected from university students.

If the quantity of data and the number of factors are raised, then reliable forecasts about student 's progress can be generated. For various data and challenges, it will necessitate adjustments and additions.

S. K. Mohamad & Z. Tasir explained [2] the application of DM in education is currently in its infancy, giving rise to educational data mining (EDM). EDM emerges as a paradigm for designing models, activities, procedures, and algorithms for analyzing educational data. EDM aims to uncover patterns and create predictions about learners' behaviors and accomplishments, as well as domain knowledge material, assessments, educational capabilities, and applications.The majority of EDM approaches have a DM profile that is supported by probability, machine learning, and static disciplines;classification is the most common task, followed by clustering.In terms of dangers, they are the impediments to the rational and formal growth of EDM that prevent, hinder, and obstruct it. As a result, EDM must address the absence of concrete and precise theory to build the foundation of how EDM operates.

C. Anuradha and T. Velmurugan examined the effectiveness of various decision tree algorithms in terms of accuracy and processing time[3]. Arpit Trivedi's work has proposed a simple method for categorizing student data using a decision tree-based technique. They compiled a database of 100 students' marks in five subjects for each of four courses. For implementing measures of a specific student's type, a frequency measure is employed is used an extracting features. To create a trained classifier, each student's most frequent five topic marks are used. They automatically predicted the class for indefinite students using a trained classifier.

The First and Second classes were discovered to have had a major impact on the classifying process. With a bigger sample dataset, the study might be expanded to look at the performance of alternative categorization techniques.

Based on certain criteria via student's selection, the rate of predictions are not consistent amongst algorithms.

V.L. Miguéis et al proposed that Within the study, data from the first year of a student's educational life (path) is utilized to propose a two-staged model that employs data mining methods to predict their eventual academic accomplishment. [4]. Unlike much educational data mining literature, educational attainment is typically determined both from average score received and the duration required to finish the degree. In addition, this study suggests dividing it up students depending on the disparity in either indications of failure or high performance just at commencement of the degree course and the model's predicted performance levels. The research evidence reveals that the proposed model can properly predict students' quality performance standards during a preliminary phase in their educational endeavors with a 95 percent overall accuracy. The approaches may run into certain technical challenges during the data extraction and training phases, causing the accuracy rate to drop by a few percentage points.

Siti Dianah et al made us understand predictive analytics used in sophisticated analytics,[5] which included machine learning deployment, to generate high-quality

performance and meaningful data for students at all levels of education. Most individuals realize that one of the critical performance factors that teachers should use to track a student 's educational development is their grades. Many different machine learning algorithms have been proposed in the education arena during the last decade. Most individuals realize that one of the critical performance factors that teachers should use to evaluate a student 's educational development is their grades.

Pendro Manuel et al states that The emergence of learning analytics has enabled the construction of predictive analysis to determine learners' behaviors and actions (e.g., performance). Several of these methods, on the other hand, are just useful in particular learning scenarios, and it can be tricky to sort to see which factors, which includes predictor variables and forecast outcome type, impact predictive findings.[6] Additional details might be helpful in making generalizations to different scenarios, comparing strategies, developing prediction models, and developing more workable alternatives. For concept-oriented assignments, predictive power was also higher, and the best models frequently just included the final encounters. Additionally, it was discovered that multiple-choice problems were better at predicting than programming ones. There are certain methodological flaws to consider. One limitation is the method for filtering students. Several trainees don't really connect with the platform and thus should be removed; additional factors might hold, and the outcomes may well be impacted.

Brijesh Kumar Baradwaj & Saurabh Pal made us realize that knowledge is buried in the educational data set, but it can be extracted using data mining techniques. By providing a data mining model for the institution's higher education sector, the focus of this research is to show the strength of data mining techniques inside the context of teaching and learning. [7] The categorization job works with a file system to anticipate student segregation based on historical data. As there are numerous methodologies for data classification, the tree based method is used here.

Elizabeth Boretz states that the common use of the word "grade inflation" in the context of higher education is a possibly harmful exaggeration.[8] Test scores are at an all-time peak, but research shows that perhaps the surge in professional learning initiatives and the increased diversity of academic advising are unrelated. Students are not customers looking for great marks in exchange for positive teacher ratings; instead, they want to succeed by acting together to assist their learning, and colleges and institutions are up to scratch.

Rathee A, Mathur RP wrote that metadata is a learning of management and data mining approach for combining collectively related resources. In the research, there seem to be a variety of classification algorithms, but decision tree algorithm are the most widely used since they are simple to write and grasp especially when compared to other algorithms. [9] The ID3, C4.5, and CART decision tree algorithms were utilized to forecast the future of students' data.

Tarun Verma et all stated that Literacy has long been a global problem. Every country strives for a 100% literacy rate. If the literacy rate has improved significantly, there is still a need to understand the areas where people are still falling behind. As a result, overall literacy statistics should be investigated in order to give a quick and correct framework for assistance inside the management and planning of educational services,[10] as well as to create or assist to a data collection, organization, and consuming platform for education data. Georgia, Cube, Estonia, Latvia, and Barbados have the highest literacy rates in the developed world, whereas Mali, South Sudan, Ethiopia, Niger, and Burkina Faso have the poorest literacy levels. When it comes to continents, the United States has the largest literacy level, followed by Europe, Asia, and Africa. Kerala does have the highest literacy rate (93.9%) in India, while Bihar has the lowest literacy rate (63.8%). In terms of area, the South has the greatest literacy rate, followed by the East, West, and North.

Abdulmohsen Alkushi & Abdulaziz Althewini proposed that even though student performance predictor can be used to forecast Saudi student academic [11] achievement, there are considerable variations between research. Based on their academic profile, a formula was created to determine student performance. *continued down*

The survey stands out from others in Saudi Arabia because it uses a bigger sample size and emphasizes on yearly student achievement. These variations have added to the complexity of current research attempts on Saudi predisposing factors of student performance prediction, and they should be considered by researchers and policy - makers alike when making decisions.

Adel M. et al proposed that from 2008-09 to 2010-11 academic years,[12] A retrospective observational analysis was conducted using information from registers of enrolled students in the three colleges. The grades mean, assessing learning rating, and standardized tests score have been the evaluation's independent variables. The outcome variable was the mean of the individuals' first- and second-year grade point averages (GPA).

It wasn't really indicative of all these participants' initial educational success in KSU's health colleges, according to the findings of this study.

Hanan Mengash suggested a methodology which was validated using data from 2,039 students enrolled in a Saudi public university's Computer Science and Information College from 2016 to 2019. The findings show that based on key parameters including high school grade average,

[13] Prior entrance, individuals' initial university success can be forecasted using their Academic Aptitude Admission Test and Associated With particular Test scores. The data also imply that a child's performance on the Academic Aptitude Tests is the best predictor of future achievement.

Mashael A. Al-Barrak and Muna Al-Razgan utilized everything in their reach to help students improve their performance and uncover early predictors of their ultimate grade point average. [14] We used a classification technique, specifically a decision tree, to predict students' final GPA based on their grades in earlier classes, which yielded rather accurate findings.

Mikko Vinni, Wilhelmiina Hämäläinen stated that intelligent tutoring systems and adaptable learning environments both require automatic classification.[15] The learner's current circumstance should be classified first. This will necessitate the use of a classifier, which is a model that's used to predict the category worth depends on other explaining criteria. Teaching method can benefit from such expectations as well, but computers learning environments can handle larger class sizes and gather more information for classifier improvement.

Zlatko J. Kovačić gave an approach in which the [16] transfer literature has identified scholastic and cultural variables (age, sex, ethnic origin, marital status) as potential predictive variables of attrition. Students' registration forms include the only knowledge, or factors, we know about them at the time of signing at the Open Polytechnic of New Zealand.The issue we are attempting to answer in this work is whether we can predict study outcomes for freshly enrolled students only based on enrollment data. The total categorization accuracy of the CHAID tree was 59.4 percent, while the CART tree was slightly higher at 60.5 percent, both of which are the highest accuracy percentages ever obtained.

Ashutosh Nandeshwar, Subodh Chaudhari used student admissions data to build models that forecast enrollment, and to evaluate the models using cross-validation, win-loss tables, and quartile charts. [17] Financial aid was employed as a controlling factor regardless of their high school GPA and ACT/SAT results. This component proved to be the most important in terms of improving the quality of new students. To see the influence of attributes like distance from campus and initial point of contact, they need to be built. The enrollment indicator and the "persistence indicator" should be combined to uncover attributes that affect retention.

Diego Garc´ıa-Saiz, Marta Zorrilla proposed that the need for teachers to forecast their students' performance is increasing as [18] virtual teaching becomes more popular. Different machine learning approaches can be utilized to address this need. In this research, we evaluated the effectiveness and perception of multiple classification techniques have been applied to schooling datasets, and we propose a meta-algorithm to modify the sources of data and improve the accurateness. The meta-algorithm given, which uses both Nave Bayes and J48 to analyze and forecast, produces the best outcomes, with the precompiled task being completed as of the most important attribute for the preprocessing technique being superior.

Pauziah Mohd Arsad et al said that report discusses advancements in forecasting engineering students' academic achievement. At the end of semester eight, the academic [19] achievement was measured using the cumulative grade point average (CGPA). The research was carried out at the Universiti Teknologi MARA (UiTM) Faculty of Electrical Engineering in Malaysia. This model is limited to electrical degree students at the Faculty, but it can be extended to other departments with the addition of appropriate input variables. For each student's software, the designer must determine the input predictor variables.

Cristóbal Romero et al said that The aim of using internet discussion discussion boards is only to see how the choosing of cases and qualities, by use of various classifiers, and the schedule when data are collected impact predictive performance and conciseness is to see if making a prediction accuracy at the completion of the term and an accurate warning well before end of the program is relevant. [20] Producing better quality when just the messages relevant to the topic are used. Without grouping and connection criteria, nevertheless, we cannot obtain reliable and generally easy to interpret models.

Farshid Marbouti et al proposed that it is possible to[21] identify at-risk pupils early and notify both the teachers and the students using predictive modeling tools. But due to the unpredictability of students' behavior, it is impossible to create a 100 percent accurate model of their performance.

G.Gray, Colm McGuinness, P. Owende's study looked at psychometric factors that may be tested early after enrollment, such as personality, motivation, and learning techniques. Model accuracy was assessed via cross validation,[22] and the results were compared to outcomes when the models were used in a following academic year. When just under 21 pupils were used to train the models, they all improved in accuracy.

Ahmed Mueen et al proposed that three classifiers' prediction performance is assessed and compared. This research will assist teachers in helping students enhance their academic performance. The Nave Bayes classifier exceeds the other two by achieving an overall prediction accuracy of 86%. [23] The pupils were not interested in using the forum because there were no assigned marks for doing so.

Gökhan Akcapinar et al's analyses were carried out using the Orange data mining tool, and the models were assessed using ten-fold [24] cross-validation. The classification model correctly predicted 22 of the 27 failing students (81.5%) and 45 of the 49 passing students (91.8 percent).

In another paper, a maximum of 76 2nd undergraduate enrolled students in a Computer Hardware program participated in the study. [26] The research attempts to know two fundamental questions by assessing different classification models and pre-processing techniques: which methods and attributes largely determine students' edge school achievement, and whether academic

achievement can be anticipated in sooner weeks to use these functionalities and the chosen algorithm.

George Siemens, Ryan S J.d. Baker proposed that Two research communities have formed to address the demand: Educational Data Mining (EDM) and [25] Learning Analytics and Knowledge Management (LAK). This study supports for improved and formalized communication and cooperation across both groups in order to share information, processes, and technologies for data mining and analysis in the purpose of increasing both the LAK and EDM domains. Many modern data mining/analytics solutions, however, need not directly endorse this goal.

M.I. López et al proposed that in order to see if student engagement in the course forum is a good predictor of final course grades, and to see if the proposed clustering approach can achieve equivalent accuracy to standard classification algorithms. The EM clustering approach achieves a level of accuracy that is similar to classic [27] classification algorithms once more. For educators, manually evaluating messages is a tough and time-consuming operation.

Jiechao Cheng stated that Data mining is a strong analytical tool for enterprises to improve decision-making and analyze new patterns and relationships, and EDM [28] includes techniques such as data mining, statistics, and machine learning. However, the costs and challenges of deploying EDM applications are high.

Ryan Shaun Baker, Paul Salvador Inventado proposed that these strategies are used by researchers and practitioners to investigate new constructs and answer new research questions. A variety of tech is used in the process of data collection for data mining [29]. Every bit of data created requires its own preservation and maintenance. As a consequence, the installation cost could increase. Furthermore, an expert must be hired for tooling and other operations, which will increase the entire price.

Chong Ho Yu et al stated that an internally produced exam, according to this analysis, can provide [30] arguably more accurate information about a student's cognitive abilities. Because the data was taken from a single institution's data warehouse, the findings cannot be applied to a broader, national scale until more replication studies are completed.

Parneet Kaur et al study aids institutions in identifying pupils who are slow learners, which may then be used to determine additional [31] assistance for them. EDM is still in its infancy, but it has a lot of educational potential. The precision was attained using the Multi Layer Perception classifier, nevertheless it was only 75%.

Zahyah Alharbi et al proposed that the most important part is to identify weak pupils who are at risk of receiving a lower grade or dropping out of school. Designers recognise circuits that are affiliated with good & evil results for children with similar traits and academic ability documents when [32] device choices are accessible, because when component options are usable, those components can be proposed or disheartened for students of similar character traits and academic ability records. In this case, the solution may be a decision support system that evaluates the paths and successes of like students to advise what could have been the top choices for a specific student. We may also look at how module dependencies and relationships are measured.

In a paper, E. Smith and P. White suggested to investigate the qualities of candidates to multiple disciplines, as well as to function each topic plays in influencing the likelihood of finishing with a 'outstanding' gpa Despite major and within variance in graduate results, multiple regression analysis between educators' social and academic features and university accomplishment disclosed that the topic educators researched had some predictive power in predicting future final degree categories upon attempting to control for cultural class and previous attainment. This finding has ramifications for main element at increasing the number[33] and level of STEM grads in a profession that is frequently known to as a "shortage" or "priority."

D. KABAKCHIEVA,K. STEFANOVA, V. KISIMOV's work talks about specific goal is to uncover interesting patterns in the [35] available data that can help forecast student performance at university based on their personal and pre-university characteristics. The goal of the data mining project is to forecast student university performance based on personal and pre-university factors.

Olugbenga Adejo and Thomas Connolly The concept behind this framework is to use a holistic strategy to accurately and efficiently anticipate student performance. The six variable domains that have a significant impact on student performance will be used in the performance prediction framework provided. [34]

• Self-efficacy, achievement, goal, and interest are examples of psychological domains.

• Exam score, presenting ability, and intellectual competence are all part of the cognitive domain.

• Motivation, learning style, study time, habit, ICT skill, and online activities are all part of the personality domain.

• Income, income distribution status, parent financial position, and employment status are all part of the economic sphere.

• Age, gender, location, ethnicity, marital status, and disability are all demographic domains. • Course programme, learning environment, institutional support, and course workload are all part of the institutional domain.

Liang Zhao et al conducted an experiment which is undertaken based on an actual university database of college kids, which combines multisource behavioral data from multiple sources, including not just physical and digital education, but also within even outside the behaviors in the classroom Performance measures assessing linear and nonlinear behaviors (e.g., regularity

and stability) of school grounds ways of life are approximated in [36] to obtain in-depth understanding through into functionalities that ultimately led to superb or underperformance; likewise, includes able to represent variability in chronological lifestyle patterns are derived utilizing long poor memory (LSTM). (2) Secondly, machine learning-based classifications approaches are being proposed to estimate academic achievement. (3) Lastly, tangible advice is being produced to assist the students (especially at-risk students) in improving their university connections and achieve a good study-life mix.

3. CONCLUSION

In this paper, we went through different methodologies, techniques and ways in which the Student Prediction Model can be made, trained & tested for the accuracy. Some statistical and neurological factors were also taken into consideration and using that it helped in carving a path of combining few ideas or modifying the existing one for a better method.

By far, through reading and understanding the case studies and the research papers, we can possibly conclude that decision tree method is utilised in this because there are several ways to data classification. To anticipate the performance of the student at the end of term, information such as attendance, class test, seminar, and assignment marks were gathered from the student's previous database.

ACKNOWLEDGEMENTS

REFERENCES

[1] Guruler, H., Istanbullu, A. & Karahasan, M. (2010). A new student performance analyzing system using knowledge discovery in higher educational databases. Computers & Education, 55(1), 247-254. Elsevier Ltd. Retrieved January 17, 2022 from https://www.learntechlib.org/p/66621/.

[2] S. K. Mohamad and Z. Tasir, ``Educational data mining: A review.'' Procedia Social Behav. Sci., vol. 97, pp. 320324, Nov. 2013.

[3] Anuradha, C & Thambusamy, Velmurugan. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. Indian Journal of Science and technology. 8. 974-6846. 10.17485/ijst/2015/v8i15/74555.

[4] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, ``Early segmentation of students according to their academic performance: A predictive modeling approach,'' Decis. Support Syst., vol. 115, pp. 3651, Nov. 2018.

[5] Bujang, Siti & Selamat, Ali & Ibrahim, Roliana & Krejcar, Ondrej & Herrera-Viedma, Enrique & Fujita, Hamido & Ghani, Nor. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3093563.

[6] Moreno-Marcos, Pedro & Pong, Ting-Chuen & Merino, Pedro & Delgado-Kloos, Carlos. (2020). Analysis of the Factors Influencing Learners' Performance Prediction With Learning Analytics. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2963503.

[7] Baradwaj, Brijesh & Pal, Saurabh. (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Applications. 2. 63-69. 10.14569/IJACSA.2011.020609.

[8] Boretz, Elizabeth. (2004). Grade Inflation And The Myth Of Student Consumerism. College Teaching. 52. 42-46. 10.3200/CTCH.52.2.42-46.

[9] A. Rathee, R. Mathur (2013) Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance(Published in BIOINFORMATICS 8 March 2013)

[10] Tarun Verma, Sweety Raj, Mohammad Asif Khan, Palak Modi (2012) Literacy Rate Analysis | The research paper published by IJSER journal is about Literacy Rate Analysis 1 | ISSN 2229-5518

[11] Alkushi, Abdulmohsen & Althewini, Abdulaziz. (2020). The Predictive Validity of Admission Criteria for College Assignment in Saudi Universities: King Saud bin Abdulaziz University for Health Sciences Experience. International Education Studies. 13. 141. 10.5539/ies.v13n4p141.

[12] Ability of admissions criteria to predict early academic performance among students of health science colleges at King Saud University, Saudi Arabia by Adel M. Alhadlaq, PhD; Osama F. Alshammari, BDS; Saleh M. Alsager, PhD; Khalid A. Fouda Neel, MD; Ashry G. Mohamed, DrPh (2015) Ability of admissions criteria to predict early academic performance among students of health science colleges at King Saud University, Saudi Arabia, J Dent Educ. 2015 Jun;79(6):665-70. PMID: 26034031

[13] Mengash, Hanan. (2020). Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2981905.

[14] Al-Barrak, Mashael & Al-Razgan, Muna. (2016). Predicting Students Final GPA Using Decision Trees: A

Case Study. International Journal of Information and Education Technology. 6. 528-533. 10.7763/IJIET.2016.V6.745.

[15] Wilhelmiina Hämäläinen & Vinni, Mikko. (2010). Classifiers for Educational Data Mining. 10.1201/b10274-7.

[16] Kovacic, Zlatko. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. 647-665. 10.28945/1281.

[17] Nandeshwar, Ashutosh & Chaudhari, Subodh. (2009). Enrollment Prediction Models Using Data Mining.

[18] García-Saiz, Diego & Zorrilla, Marta. (2011). Comparing classification methods for predicting distance students' performance. Journal of Machine Learning Research - Proceedings Track. 17. 26-32.

[19] Mohd Arsad, Pauziah & Buniyamin, Norlida & Ab Manan, Jamalul-Lail. (2013). A neural network students' performance prediction model (NNSPPM). 1-5. 10.1109/ICSIMA.2013.6717966.

[20] Romero, Cristóbal & López, Manuel-Ignacio & Luna, José María & Ventura, Sebastian. (2013). Predicting Students' Final Performance from Participation in On-line Discussion Forums. Computer Education. 68. 458-472. 10.1016/j.compedu.2013.06.009.

[21] Marbouti, Farshid & Diefes-Dux, Heidi & Madhavan, Krishna. (2016). Models for early prediction of at-risk students in a course using standards-based grading. Computers & Education. 103. 10.1016/j.compedu.2016.09.005.

[22] Gray, Geraldine & McGuinness, Colm & Owende, Philip. (2014). An application of classification models to predict learner progression in tertiary education. Souvenir of the 2014 IEEE International Advance Computing Conference, IACC 2014. 549-554. 10.1109/IAdCC.2014.6779384.

[23] Mueen, Ahmed & Zafar, Bassam & Manzoor, Umar. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International Journal of Modern Education and Computer Science. 11. 36-42. 10.5815/ijmecs.2016.11.05.

[24] Gökhan Akcapinar, Arif Altun, Petek Askar(6 August 2015)Modeling Students' Academic Performance Based on Their Interactions in an Online Learning Environment, SOSED Holistic Education Consultancy & Publications

[25] Siemens, George & Baker, Ryan. (2012). Learning analytics and educational data mining: Towards communication and collaboration. ACM International Conference Proceeding Series. 10.1145/2330601.2330661.

[26] Akçapınar, G., Altun, A. & Aşkar, P. Using learning analytics to develop early-warning system for at-risk students. Int J Educ Technol High Educ 16, 40 (2019). https://doi.org/10.1186/s41239-019-0172-z

[27] Lopez, M.I. & Luna, José María & Romero, Cristóbal & Ventura, Sebastian. (2012). Classification via clustering for predicting final marks based on student participation in forums. Proc. of 5th Int. Conf. on Educational Datamining. 148-151.

[28] Cheng, Jiechao. (2017). Data-Mining Research in Education.

[29] Baker, Ryan & Inventado, Paul. (2014). Educational Data Mining and Learning Analytics. 10.1007/978-1-4614-3305-7_4.

[30] Yu, Chong Ho & Yu, Samuel & Digangi, Angel & Jannasch-Pennell, Charles & Kaprolet,. (2010). A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year. Journal of Data Science. 8. 307-325. 10.6339/JDS.2010.08(2).574.

[31] Kaur, Parneet & Singh, Manpreet & Josan, Gurpreet. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science. 57. 500-508. 10.1016/j.procs.2015.07.372.

[32] Alharbi, Zahyah & Cornford, James & Dolder, Liam & Iglesia, Beatriz. (2016). Using data mining techniques to predict students at risk of poor performance. 523-531. 10.1109/SAI.2016.7556030.

[33] Smith, Emma & White, Patrick. (2014). What makes a successful undergraduate? The relationship between student characteristics, degree subject and academic success at university. British Educational Research Journal. 41. 10.1002/berj.3158.

[34] Adejo, Olugbenga & Connolly, Thomas. (2017). An Integrated System Framework for Predicting Students' Academic Performance in Higher Educational Institutions. International Journal of Computer Science and Information Technology. 9. 149-157. 10.5121/ijcsit.2017.93013.

[35] Kabakchieva, Dorina & Stefanova, Kamelia & Kisimov, Valentin. (2011). Analyzing University Data for Determining Student Profiles and Predicting Performance.. 347-348.

[36] L. Zhao et al., "Academic Performance Prediction Based on Multisource, Multi Feature Behavioral Data," in IEEE Access, vol. 9, pp. 5453-5465, 2021, doi: 10.1109/ACCESS.2020.3002791.