INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

Attribute Prediction for Undergraduate Students by Applying Machine Learning Techniques

Mr. S. Sooryanath, Dr. S. Anguraj M.E, Ph.D

Department of Information Technology, KSR College of Engineering, Tiruchengode, India

ABSTRACT: This study presents six machine learning models in the prediction of student success in a technology-mediated environment. Student behavioral attributes with a learning management environment have proven to be a significant determinant in forecasting students' performance. This study attempts to provide the model with optimum accuracy to determine students who need assistance to improve their educational performances and other learning outcomes. Attribute Selection with the top ten attributes and 10-fold cross-validation offer best performances. The six predictive models utilized in this study are Linear Discriminant Analysis, Logistic Regression, Classification and Regression Tree, K-Nearest Neighbour, Naïve Bayes Classifier, and Support Vector Machines. Classification and Regression Tree model and Linear Regression had the best accuracy score of 0.86 after 10-fold crossvalidation and top ten attribute selection. This study concludes that student behavioral attributes are useful predictors of student success.

I. INTRODUCTION

The machine learning technique is one of the main methods used in studying student performance or success, aside from statistical analysis and data mining. Academic performance is a daunting challenge for tertiary education institutions across the globe. [1] and [2] described

data analytics as a tool for identifying students who are struggling educationally and enhancing throughput in various educational institutions. This study falls under the category of Educational Data Mining (EDM). EDM is a subdivision of data mining that specializes in designing, evaluating, and implementing various automatic tools for measuring vast amounts of data from academic environments [3]. This study investigates student success through student behavioral attributes in the learning management environment. In any learning environment, student engagement is a crucial indicator for assessing a student's success or failure [4]. The channels of education delivery include traditional classroom, online-learning, blended-learning, and others. The Learning Management System (LMS) is a learning platform that allows instructors and learners to communicate without having to meet in person [5]. The global adoption of LMS platforms in learning is increasing by the day as several factors have warranted this acceptance. The reason for the adoption of LMS include the convenience of learning at student pace, improvement of cost-efficiency for the institutions and full coverage of a large number of students [6]. The blended-learning, which is interchangeably called hybrid-learning, is an infusion of both standard classroom and technology-aided settings [7]. [8] provided objectives of blended-learning as an effective learning process, studentteacher physical contact, academic performance enhancement, and learner's freedom.

II. LITERATURE SURVEY

1. Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques
Author- Fang-Yie Leu
Contribution: This study applies supervised and unsupervised machine learning (ML) techniques to discover which significant attributes that a successful learner often demonstrated in a computer course. Background: Students often experienced difficulties in learning an introduction to computers course. This research attempts to investigate how successful students regulate their learnings in this course. The answer to these questions will provide teachers with useful information to better comprehend how students learn and which strategies are effective in learning. Research Questions: 1) Which algorithm in supervised learning is the best one for predicting students' final performance? and 2) What attributes are key to succeed in this course? Methodology: Seven supervised ML algorithms and ensembles are conducted to compare the performance of classifiers regarding the levels of accuracy, precision, and sensitivity.

2. Attributes selection using machine learning for analysing students' dropping out of university: a case study

Author - Citation T I Pehlivanova and V I Nedeva 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1031 012055

Many students in Bulgarian universities drop out of the university before completing their studies. Identifying students at risk of dropping out allows timely taking measures for their retention. The paper presents the results of a study conducted among students of engineering programs at Trakia University - Stara Zagora. The collected data are subjected to processing, which aims to find the most important attributes that determine the risk of dropping out of university. The processing is done with Weka open source software. Different algorithms for selecting attributes with different search methods are applied. The most appropriate attribute selection algorithm was selected after applying the BayesNet classifier to the results obtained. The indicators TP rate, Precision and F-measure were compared. When applying InfoGainAttributeEval, the highest results are obtained for the accuracy of the classification. At the next stage, it is planned to expand the study among a larger number of students from different programs and create an effective forecasting model.

III. EXISTING SYSTEM

We propose a web based intelligent student advising system using collaborative filtering, a technique commonly used in recommendation systems assuming that users with similar characteristics and behaviors will have similar preferences. With our advising system, students are sorted into groups and given advice based on their similarities to the groups. If a student is determined to be similar to a group student, a course preferred by that group might be recommended to the student.

Disadvantages

- ➢ System used to predict suitable course for students and data-set not compatible to predict student results.
- ➢ Not all student behaviors connected to course advising.
- ➢ Students are grouped and then system predicts the suitable course for the students. Grouping lacks over data for prediction.

PROPOSED SYSTEM

The major objective is to find behaviour patterns of students in a timely and accurate manner. Main aim is specifically to identify the categories of students who require extra attention. Without campus behaviour analysis, These students' academics and several other performance dimensions are impacted. System uses parameters such as attendance status, extra circular activities, grade, technical skills, previous semester results, grasping capability, Aptitude grade, interaction with lecturers etc.. System helps lecturers to identify the most influential factors affecting the students performance. System uses data science technique called as "Association Learning" to find the patterns. We use "Eclat algorithm" to find patterns and classification algorithm i.e "Naïve bayes Algorithm" to make individual predicition of student's performance. Proposed system to build as web application useful to colleges and lecturers to know the students behaviour patterns.

DISADVANTAGES
- ➢ Proposed project is a studentsbehavior analysis and prediction and management system which is meant for educational institute.
- ➢ Proposed project makes use of ML technique or Data Science approach for the studentsbehavior analysis.
- ➢ To provide valid information from existing students to manage relationships with upcoming students.
- ➢ Identification of different factors which affects a student's learning behavior and performance during academic career.
- ➢ Construction of a prediction model using ML technique - Unsupervised Learning on the bases of identified predictive variables.

IV. EXPERIMENTAL RESULT

In current system it is difficult to track the student's behavior and characteristics. There is no automation or tool which predicts or shows how to improvise the student's academic performances. Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. Machine learning (ML) algorithms is use to better understand learners' learning is popular in the educational community. This paper presents examining successful attributes prediction for undergraduate students by machine learning techniques.
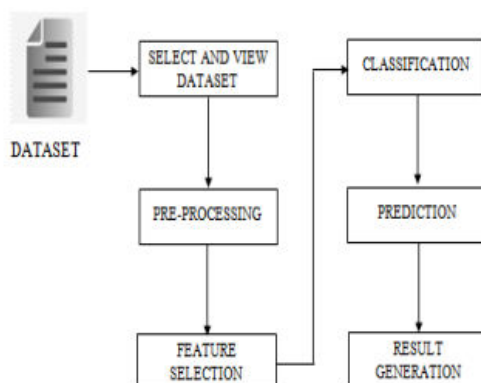
Figure 2: Flow Chart

Data Selection and Loading-
The data selection is the process of selecting the data for Student Grade dataset.

Data Preprocessing
- Data pre-processing is the process of removing the unwanted data from the dataset.
- Missing data removal
- Encoding Categorical data

Splitting Dataset into Train and Test Data
- Data splitting is the act of partitioning available data into two portions, usually for cross-validate purposes.
- One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.

V. CONCLUSION

The present study proposes a prediction model and identifies critical attributes (i.e., self-efficacy beliefs and to make sure keep up with the weekly progress in the class) for success in the class. Based on the findings, teachers can consider how to provide an appropriate learning environment to support students to be successful learners. One possible solution is that an e-assessment for students to self-evaluate their learnings anytime and anywhere weekly. The immediate feedback from self-assessment allows students to examine their understandings in a timely manner; the sustainable learning purpose is obvious. Teachers can also gather and analyze learners' assessment records to diagnose their learning problems. In addition, through regular online self-assessments, stude have the opportunities to enhance their self-efficacy beliefs and further, to improve their learnings ultimately. Although the background of this study is in a computer course, the methodology proposed by this research can still apply to other courses or different grade levels. Researchers can adjust the composition of students' final grades based on the nature of the course. Hence, researchers need to consider the impact of these factors on student achievement before building their learning models. This study demonstrates the effectiveness of employing ML algorithms to discover the critical attributes of successful learners.

Future Enhancements
This study presents six machine learning models in the prediction of student success in a technology-mediated environment. Student behavioral attributes with a learning management environment have proven to be a significant determinant in forecasting students' performance. This study attempts to provide the model with optimum accuracy to determine students who need assistance to improve their educational performances and other learning outcomes. We examined the impacts of SMOTE data re-sampling and the effect of attribute selection in this study. The models' performances were enhanced with the resampling method as the imbalanced dataset was identified to have performed poorly. Attribute Selection with the top ten attributes and 10-fold cross-validation offer best performances. The six predictive models utilized in this study are Linear Discriminant Analysis, Logistic Regression, Classification and Regression Tree, K-Nearest Neighbour, Naïve Bayes Classifier, and Support Vector Machines. Classification and Regression Tree model and Linear Regression had the best accuracy score of 0.86 after 10-fold crossvalidation and top ten attribute selection. This study concludes that student behavioral attributes are useful predictors of student success.

REFERENCES

[1] R. Ajoodha, A. Jadhav, and S. Dukhan, "Forecasting learner attrition for student success at a south african university," in In Conference of the South African Institute of Computer Scientists and Information Technologists 2020 (SAICSIT '20), September 14-16, 2020, Cape Town, South Africa. ACM, New York, NY, USA, 10 pages. ACM, 2020.

[2] A. D. Kumar, R. P. Selvam, and K. S. Kumar, "Review on prediction algorithms in educational data mining," International Journal of Pure and Applied Mathematics, vol. 118, no. 8, pp. 531–537, 2018.

[3] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert systems with applications, vol. 33, no. 1, pp. 135–146, 2007.

[4] M. Hu and H. Li, "Student engagement in online learning: A review," in 2017 International Symposium on Educational Technology (ISET). IEEE, 2017, pp. 39–43.

[5] A. Dutt and M. A. Ismail, "Can we predict student learning performance from lms data? a classification approach," in 3rd International Conference on Current Issues in Education (ICCIE 2018). Atlantis Press, 2019, pp. 24–29.

[6] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student engagement predictions in an e-learning system and their impact on student course assessment scores," Computational intelligence and neuroscience, vol. 2018, 2018.

[7] W. W. Porter, C. R. Graham, K. A. Spring, and K. R. Welch, "Blended learning in higher education: Institutional adoption and implementation," Computers & Education, vol. 75, pp. 185–195, 2014.

[8] R. T. Osguthorpe and C. R. Graham, "Blended learning environments: Definitions and directions," Quarterly review of distance education, vol. 4, no. 3, pp. 227–33, 2003.

[9] T. Abed, R. Ajoodha, and A. Jadhav, "A prediction model to improve student placement at a south african higher education institution," in 2020 International SAUPEC/RobMech/PRASA Conference. IEEE, 2020, pp. 1–6.

[10] Z. Cai, X. Fan, and J. Du, "Gender and attitudes toward technology use: A meta-analysis," Computers & Education, vol. 105, pp. 1–13, 2017.

[11] C. J. Asarta and J. R. Schmidt, "Comparing student performance in blended and traditional courses: Does prior academic achievement matter?" The Internet and Higher Education, vol. 32, pp. 29–38, 2017.

[12] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," International Journal of Database Theory and Application, vol. 9, no. 8, pp. 119–136, 2016.

[13] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student performance prediction model based on supervised machine learning algorithms," in IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3. IOP Publishing, 2020, p. 032019.

[14] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from vle big data using deep learning models," Computers in Human Behavior, vol. 104, p. 106189, 2020.

[15] K. Mongkhonvanit, K. Kanopka, and D. Lang, "Deep knowledge tracing and engagement with moocs," in Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 2019, pp. 340–342.

[16] J. C. S. Silva, J. L. Ramos, R. L. Rodrigues, A. S. Gomes, F. d. F. de Souza, and A. M. A. Maciel, "An edm approach to the analysis of students' engagement in online courses from constructs of the transactional distance," in 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT). IEEE, 2016, pp. 230–231.

[17] J. Gardner and C. Brooks, "Student success prediction in moocs," User Modeling and User-Adapted Interaction, vol. 28, no. 2, pp. 127–203, 2018.

[18] T. Soffer and A. Cohen, "Students' engagement characteristics predict success and completion of online courses," Journal of Computer Assisted Learning, vol. 35, no. 3, pp. 378–389, 2019.

[19] A. Kamath, A. Biswas, and V. Balasubramanian, "A crowdsourced approach to student engagement recognition in e-learning environments," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016, pp. 1–9.

[20] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Relationship between student engagement and performance in e-learning environment using association rules," in 2018 IEEE World Engineering Education Conference (EDUNINE). IEEE, 2018, pp. 1–6.

[21] M. A. Alsubhi, N. S. Ashaari, and T. S. M. T. Wook, "The challenge of increasing student engagement in e-learning platforms," in 2019 International Conference on Electrical Engineering and Informatics (ICEEI). IEEE, 2019, pp. 266–271.

[22] R. Longadge and S. Dongre, "Class imbalance problem in data mining review," arXiv preprint arXiv:1305.1707, 2013.

[23] R. Ajoodha, S. Dukhan, and A. Jadhav, "Data-driven student support for academic success by developing student skill profiles," in 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC). IEEE, 2020, pp. 1–8.

[24] J. Brownlee, "Machine learning mastery with python," Machine Learning Mastery Pty Ltd, pp. 100–120, 2016.

[25] S. Sperandei, "Understanding logistic regression analysis," Biochemia medica: Biochemia medica, vol. 24, no. 1, pp. 12–18, 2014.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 **9940 572 462** 🟢 **6381 907 438** ✉ **ijircce@gmail.com**