



Examining Successful Attributes Prediction for Undergraduate Students by Machine Learning Techniques

Mohd Tabish Idrishi¹, Sonal Chaudhary²

M.Tech Scholar¹, Associate Professor², Department of Computer Science & Engineering, All Saints' College of Technology, Bhopal, Madhya Pradesh, India

Abstract— 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.

Keywords— Undergraduate Students, Examining Successful Prediction, Machine Learning, AI.

I. INTRODUCTION

Consistently, a huge number of understudies fail basic programming courses around the world, and various understudies pass their courses with inadequate information. As a con-grouping, studies are retaken and deferred, professions are revaluated, and significant capital is put into understudy guiding and support. Around the world, on normal 33% of understudies fail their initial programming course. In any event, while seeing measurements portraying pass rates in the wake of showing mediations, upwards of one fourth of the understudies actually fail the courses. One of the difficulties in sorting out showing mediations is that any change is probably going to likewise affect understudies for whom the predominant circumstance is more appropriate. For instance, in the event that an understudy is as of now at a phase where she could work on additional difficult undertakings all alone, obligatory exorbitantly organized learning exercises that everybody needs to follow may try and be counterproductive for her. To give another model, while cooperative learning practices, for example, match programming have been featured as efficient showing approaches for basic programming, there are settings in which understudies generally work from a good ways and seldom go to an establishment.

Significant data with respect to understudy learning and work on the understanding of the finding, the current examination embraces the self-guideline hypothesis as a structure and utilizes both managed and solo learning strategies to investigate which credits are basic for an understudy during his/her learning. A significant issue for understudy undertakings the board is the logical inconsistency between the restricted energy of understudy guides and the variety of understudy ways of behaving, which brings about numerous potential issue understudies losing the chance for early intercession. Starting from the start of the 21st 100 years, the quick improvement of data innovation in training and the development of computerized grounds has made it feasible for understudy advocates to lead quantitative examination of understudy school ways of behaving, particularly to give early advance notice to understudies who might have issues, so the logical inconsistency could be reduced by applying the investigation and early admonition techniques. As contemporary understudies who experienced childhood in the Web time, their regular routine, learning and believing are profoundly affected by the Web. This furnishes us with the likelihood to comprehend their grounds network social qualities through enormous information. The most effective method to dig valuable data for understudy guide from huge information in the unstable development of information classifications and information scales, is a test for current understudy guide, likewise a significant chance to lead work by new means. Over the course of the present life, profound prosperity is a significant viewpoint to be investigated especially for school going youngsters. While possibly not appropriately broke down, the personal hardships might upset scholar, individual and social development, bringing about a long period of challenges for such people. Just when a youngster is sincerely secure and content, he/she does everything they can for each challenge they face. Schools ought to establish a gradual learning climate where scholarly accomplishment is connected not exclusively to effective learning methodologies, yet additionally to great mental prosperity [4].

II. BACKGROUND

C. Y. Ko et al., research endeavors to examine how fruitful undergraduates manage their learnings in this course. The response to these inquiries will give educators helpful data to more readily understand how undergraduates learn and which procedures are successful in learning. The action in exactness and awareness of this classifier accomplishes 83.26% and 92.88%, separately [1]. P. R. Tlalpan et al., got a p-worth of 0.0014 in the non-parametric Wilcoxon test, showing a measurably massive change in gained learning. AR exercises were executed utilizing different instructive methodologies. We reasoned that AR is a superb innovative device to help far off science classes, changing how undergraduates connect [2].

J. M. Campos et al., article portrays an information science challenge-based growth opportunity acquainted with non-IT second-year designing undergraduates. The technique proposed in this article was fruitful in the short starting course [3]. M. Varsha, et al., plan to investigate the importance, circumstance, and gatherings related with bliss in kids and to convey the best device to identify the condition of the youngsters' psychological well-being in regards to whether they are cheerful or miserable and in view of AI procedures [4].

H. Yang et al., use both explorative factual examination and information mining/AI ways to deal with first figure out how individual courses and the endorsed course groupings impact an undergraduate's dropout/graduation status, and afterward fabricate AI models to decipher/approve the noticed interdependency among key courses in the ongoing educational program [5]. S. Subha et al., work targets looking at the exactness acquired by various AI (ML) calculations for the given informational collection. Disarray framework has been produced utilizing AI calculations specifically Backing Vector Machines (SVM), k-Closest Neighbors (kNN), Strategic relapse (LR), and Irregular Backwoods [6].

M. Saifuzzaman et al., creators made a savvy SGPA and CGPA expectation project, as well as the outcomes on undergraduates. The discoveries are anticipated utilizing the Nave Byes calculation. The Nave Byes calculation is a basic yet compelling expectation calculation [7]. R. Alamri et al., research reasonable models of undergraduate execution forecast from 2015 to 2020. We break down and integrate essential examinations, and gathering them in view of nine aspects. Our examination uncovered the requirement for additional investigations on reasonable undergraduate execution expectation models, where both precision and make sense of capacity are appropriately measured and assessed [8].

L. M. Cruz Castro et al., discoveries propose that past information or openness to various practices can bring about various movements of additional complex computational works on, stressing the pertinence of investigating and troubleshooting as a training expected for a fruitful and convenient movement on the procurement of other computational reasoning practices [9]. F. Marbouti et al., presents, segment data and past scholarly records were investigated to grasp examples of undergraduate achievement. Technique: A bunch investigation was directed to comprehend gatherings of undergraduates in light of scholastic execution and segment data [10].

R. Ajoodha et al., features how a prescient model can help undergraduates, with an interest in Science to foster an expertise profile expected to find success in their undergrad Science program. This is accomplished by recognizing the contrast between the essential abilities expected to find success in a science program (determined utilizing information driven comes nearer) from the ongoing student's expertise profile (got from the students' exhibition in evaluations) [11]. D. Pašić et al., proposes one potential standard for the given test: we will foster an AI model with the gathered information from secondary schools which the undergraduates have gone to preceding Polynomial math and compute the likelihood of not completing the program effectively. Undergraduates who might be arranged by the model as fruitless will get proposal for another review program [12].

III. METHODOLOGY

The methodology or the flow of the work is as followings-

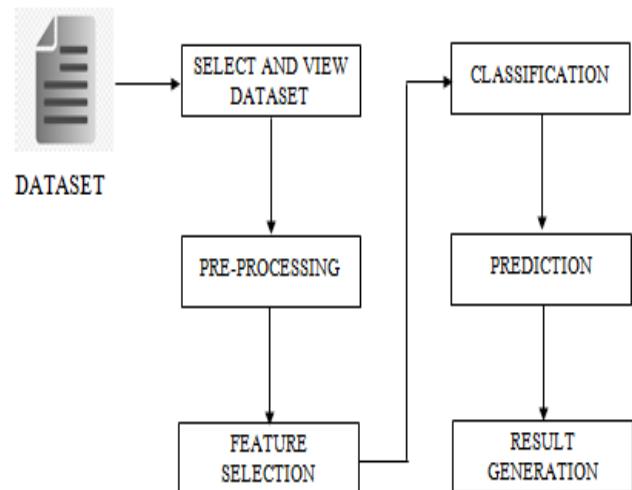


Figure 2: Flow Chart

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.

Classification

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

A multilayer perceptron is a class of feed forward artificial neural network. The term MLP is used ambiguously, sometimes loosely to mean any feed forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptron.

Prediction

- It's a process of predicting the student successful attributes from the dataset.
- This work will effectively predict the data from dataset by enhancing the performance of the overall prediction results.

IV. SIMULATION RESULTS

The execution of the proposed calculation is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library assists us with utilizing the capacities accessible in spyder climate for different strategies.

Index	school	sex	age	address	famsize
0	0	0	18	1	0
1	0	0	17	1	0
2	0	0	15	1	1
3	0	0	15	1	0
4	0	0	16	1	0
5	0	1	16	1	1
6	0	1	16	1	1
7	0	0	17	1	0
8	0	1	15	1	1
9	0	1	15	1	0
10	0	0	15	1	0
11	0	0	15	1	0
12	0	1	15	1	1
13	0	1	15	1	0

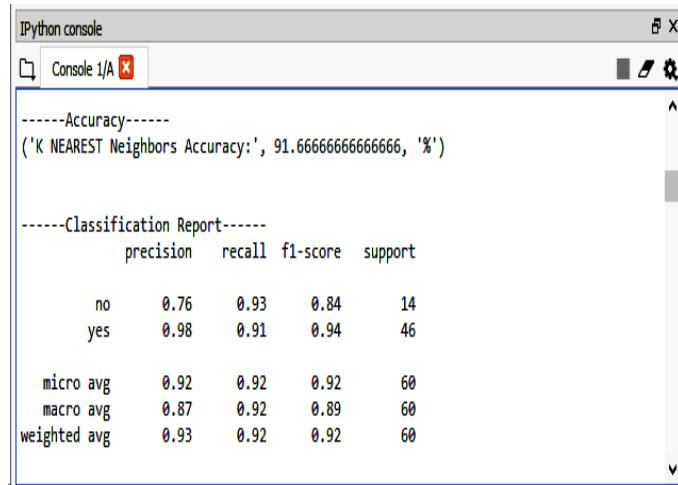
Figure 3: Dataset frame

Figure 3 is showing the dataset in the python environment. The dataset have various numbers of rows and column. The signal features name is also mentioned.

	school	sex	age	address	famsize	Pstatus	...	health	absences	G1	G2	G3	passed
0	GP	F	18	U	GT3	A	...	3	6	5	6	6	no
1	GP	F	17	U	GT3	T	...	3	4	5	5	6	no
2	GP	F	15	U	LE3	T	...	3	10	7	8	10	yes
3	GP	F	15	U	GT3	T	...	5	2	15	14	15	yes
4	GP	F	16	U	GT3	T	...	5	4	6	10	10	yes

Figure 4: Features distribution

Figure presents the various features of the dataset, the group of school, student age, address, health, marks, status of success etc.



```

IPython console
Console 1/A X
-----Accuracy-----
('K NEAREST Neighbors Accuracy:', 91.66666666666666, '%')

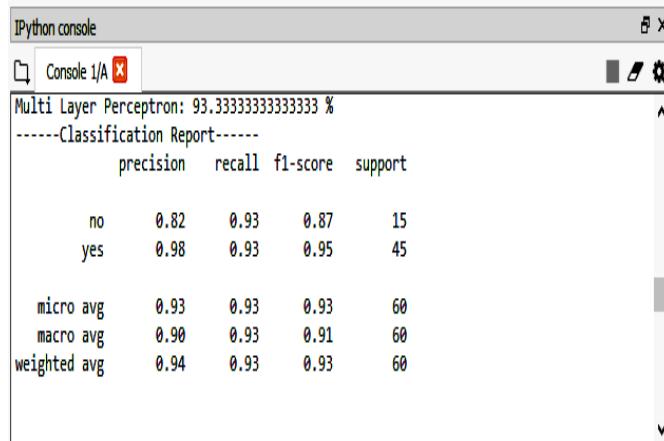
-----Classification Report-----
precision    recall    f1-score   support
no          0.76     0.93     0.84      14
yes         0.98     0.91     0.94      46

micro avg   0.92     0.92     0.92      60
macro avg   0.87     0.92     0.89      60
weighted avg 0.93     0.92     0.92      60

```

Figure 5: Simulation of the KNN technique

Figure 5 is presenting the simulation and results of the KNN technique, the overall accuracy is 91.66%.



```

IPython console
Console 1/A X
Multi Layer Perceptron: 93.33333333333333 %

-----Classification Report-----
precision    recall    f1-score   support
no          0.82     0.93     0.87      15
yes         0.98     0.93     0.95      45

micro avg   0.93     0.93     0.93      60
macro avg   0.90     0.93     0.91      60
weighted avg 0.94     0.93     0.93      60

```

Figure 6: Simulation of the MLP technique

Figure 6 is presenting the simulation and results of the MLP technique, the overall accuracy is 93.33%.

Table 5.2:
Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Method	Naïve Bayes	MLP
2	Accuracy	83.26	93.3
3	Error Rate	16.74	6.7
4	Sensitivity	92.88	97.7
5	Specificity	73.16	82.4

V. CONCLUSION

This paper presents examining successful attributes prediction for undergraduate students by machine learning techniques. Analyzing student mental issues and low academic performances is a complex task in the current education sector. System uses data. science technique called as "Association Learning" to find the patterns. The overall accuracy is achieved by the proposed technique is 93.33% while previous it is achieved by the 83.26%. The error rate is 6.7% in the proposed work and the 16.74% by the previous work. Therefore the proposed MLP machine learning technique archived better results than the previous.

REFERENCES

- [1] C. Y. Ko and F. -Y. Leu, "Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques," in IEEE Transactions on Education, vol. 64, no. 1, pp. 50-57, Feb. 2021, doi: 10.1109/TE.2020.3004596.
- [2] P. R. Tlalpan, "The Impact of Augmented Reality on Student Learning and Emotions in Mathematics," 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, 2021, pp. 1-6, doi: 10.1109/IEEECONF53024.2021.9733781.
- [3] J. M. Campos, E. A. Lozano, J. Urzúa and J. G. Calderón, "Challenge Based Learning: A Fast Track To Introduce Engineering Students To Data Science," 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, 2021, pp. 1-6, doi: 10.1109/IEEECONF53024.2021.9733765.



International Journal of Recent Development in Engineering and Technology

Website: www.ijrdet.com (ISSN 2347-6435(Online)) Volume 11, Issue 08, August 2022)

- [4] M. Varsha, M. Ramya, C. C. Sabin, N. Subheesh and J. Ali, "Assessing Emotional Well-being of Students using Machine Learning Techniques," 2021 19th OITS International Conference on Information Technology (OCIT), 2021, pp. 336-340, doi: 10.1109/OCIT53463.2021.00073.
- [5] H. Yang, T. W. Olson and A. Puder, "Analyzing Computer Science Students' Performance Data to Identify Impactful Curricular Changes," 2021 IEEE Frontiers in Education Conference (FIE), 2021, pp. 1-9, doi: 10.1109/FIE49875.2021.9637474.
- [6] S. Subha and S. B. Priya, "Comparative Analysis of Supervised Machine Learning Algorithms for Evaluating the Performance Level of Students," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 1-10, doi: 10.1109/I-SMAC52330.2021.9640798.
- [7] M. Saifuzzaman, M. Parvin, I. Jahan, N. N. Moon, F. Narin Nur and S. F. Shetu, "Machine Learning Approach to Predict SGPA and CGPA," 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), 2021, pp. 211-216, doi: 10.1109/ICAICST53116.2021.9497847.
- [8] R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," in IEEE Access, vol. 9, pp. 33132-33143, 2021, doi: 10.1109/ACCESS.2021.3061368.
- [9] L. M. Cruz Castro, A. J. Magana, K. A. Douglas and M. Boutin, "Analyzing Students' Computational Thinking Practices in a First-Year Engineering Course," in IEEE Access, vol. 9, pp. 33041-33050, 2021, doi: 10.1109/ACCESS.2021.3061277.
- [10] F. Marbouti, J. Ulas and C. -H. Wang, "Academic and Demographic Cluster Analysis of Engineering Student Success," in IEEE Transactions on Education, vol. 64, no. 3, pp. 261-266, Aug. 2021, doi: 10.1109/TE.2020.3036824.
- [11] R. Ajoodha, S. Dukhan and A. Jadhav, "Data-Driven Student Support for Academic Success by Developing Student Skill Profiles," 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2020, pp. 1-8, doi: 10.1109/IMITEC50163.2020.9334109.
- [12] D. Pašić and D. Kučak, "Machine learning model for detecting high school students as candidates for drop-out from a study program," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, pp. 1140-1144, doi: 10.23919/MIPRO48935.2020.9245405.
- [13] A. J. Fernández-García, R. Rodríguez-Echeverría, J. C. Preciado, J. M. C. Manzano and F. Sánchez-Figueredo, "Creating a Recommender System to Support Higher Education Students in the Subject Enrollment Decision," in IEEE Access, vol. 8, pp. 189069-189088, 2020, doi: 10.1109/ACCESS.2020.3031572.
- [14] V. Lara-Prieto, E. J. Arrambide-Leal, J. d. I. Cruz-Hinojosa, M. I. Ruiz-Cantísan, J. R. Rivas-Pimentel and J. Membrillo-Hernández, "Building Resilience in Engineering Students: Rube Goldberg Machine Massive Challenge," 2020 IEEE Global Engineering Education Conference (EDUCON), 2020, pp. 943-947, doi: 10.1109/EDUCON45650.2020.9125387.