



## A Machine Learning Techniques Used For Students' Academic Success Prediction

**Dr. Balasaheb B. Kalhapure<sup>1\*</sup>**

<sup>1\*</sup>Head of Dept. in Commerce, Dr. Babasaheb Ambedkar College Aundh, Pune.  
Email: kalhapurebb71@gmail.com

**\*Corresponding Author: Dr. Balasaheb B. Kalhapure**

\*Head of Dept. in Commerce, Dr. Babasaheb Ambedkar College Aundh, Pune.  
Email: kalhapurebb71@gmail.com

	<i>Abstract</i>
<b>CC License</b> CC-BY-NC-SA 4.0	<p>Young generation of every country is the future of the country. The country with the highest GER in higher education will be more successful in all the terms (Keller, K.R.I., 2021). India's GER in higher education in 2018-19 was 26.3, and in 2019-20 is 27.1. It is observed from statistics that it is which is increased. Students are enrolling for higher education but many fails to complete it (Ministry of Education, Government of India, AISHE Report 2019-20). This leads to the need of identification of reasons of students' academic success or failure. If we predict students' academic success or failure at the initial stages of their graduation period will help to take preventive measures and increase passing percentage. Student academic success is one of the criteria for accessing quality of the educational institutions, and it is one of the crucial components. There are different aspects of students' academic success, such as exam-oriented, employment-oriented, and higher study-oriented.</p> <p><b>Keywords: Machine Learning, Students, Academic Success.</b></p>

### 1. Introduction

According to Sustainable Development Goal (SDG), any educational institution's responsibility is to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (Nhamo & Mjimba, n.d.). The institute needs to identify weak students and take appropriate action to improve their performances. In this process, predicting the performance of students throughout the educational period is very important. There are students who have started with higher studies but could not complete it. So, it is necessary to find out the reason behind it. If we monitor the performance of students in the early stage, it will be beneficial for students as well as for Teachers also to follow different practices to improve the student performance.

A system or a model that could predict student success in the first few semesters of a program would substantially help to alleviate this problem. And there is a need to develop an early Student Academic Success Prediction (SASP) model to address this problem. To build such a system or model would require the identification of early performance predictors of student success and the appropriate implementation of a scientifically sound, robust modeling technique.

Early SASP is going to help students to identify their strengths and weaknesses. It will allow students to take preventive measures for completing a degree successfully. Also, in this study, the correlation between other co-curricular activities and work environment and success identification is there. The study also helps parents to find their child's progress during higher educational studies. The research aims to find the relationship between family background and student academic success, allowing parents to motivate their children and overcome family background difficulties. Family support is essential in Student' education as well as academic success(Roksa & Kinsley, 2018).

Early prediction of students' potential can be a helpful strategy to mitigate failure, achieve better outcomes, and skillful management of resources in higher education institutions (Miguéis et al., 2018). The proposed work benefits teachers who can design teaching pedagogy and takes preventive measures. The study will also help in enhancing teaching outcomes (Zheng L. et al., 2018).

Student retention is a significant challenge for different educational institutions (Aljohani, 2016). The performance of an institution is measured by various factors such as quality education, campus placement, extra-curricular activities, infrastructure, and resources. This study will help institutions develop educational policies, appoint highly qualified faculties, and improve quality education by providing the best infrastructure and resources. Skill-based education plays a vital role in employment, and it is the university's responsibility to design a course curriculum with the industrial collaboration that provides extra skill sets and fulfills industrial requirements(Borah et al., 2021). The aim of the study is to find a correlation between the socio-economic background and success of the student, which will help universities and governments to design different educational policies, program fees structure, scholarships, and financial aids programs or grants.

## 2. Literature Review

The literature reviews is essential for gaining an understanding of the existing research and explore a particular topic or area of study, and to present that knowledge. The databases such as Science Direct, IEEE Xplore, Springer, Google Scholar ,and Research Gate were referred using Machine learning ,predicting students' academic performance, academic success, student success prediction, higher education as keywords.

## 2. Objective

- i.) To design a state-of-the-art prediction model for student's academic success based on identified indicators.
- ii.) To evaluate the performance of prediction model with existing models.
- iii) To identify different indicators affecting students' academic success

## 4. Scope and Limitation

The study primarily focuses on prediction of student academic success in the Computer Science domain using different Machine Learning Techniques. This study will also design a decision support system for different stakeholders. The data of colleges under Savitribai Phule Pune University in urban area are considered for research study. There are different colleges are there under SPPU.

## 5. Research Methodology

There are take some colleges in Pune University region for the study 10% to 20% of total college data will be collected. Last 5 years up to the semester 3 or 4 data will be collected for study. The data needs for research will be broadly classified into three categories.

- i) Demographic Data: Demographic and socio-economic characteristics of individual student.
  - ii) Academic Data: Prior academic data of students like SSC score, HSC score and university data like semester marks.
  - iii) Organizational Data: data related to organization, origin, function, objectives, performance and growth.
- For the specified research primary and secondary data is needed. Secondary data will be collected directly from the institutions. Primary data will be collected fresh from student as well as educationist using different techniques like survey techniques, questionnaire

## 6. Data Collection and Analysis

Education is attaining new knowledge, understanding, skills, values, and categorized as formal or informal. "Education contributes to improved livelihoods, rapid economic development, reductions in gender inequities, strengthened support for democracy, and greater concern for the environment among other contributions" (Nhamo & Mjimba, n.d.).In India formal education is categorized into four levels: secondary, higher secondary, graduation, and post-graduation level. Indian Institutions are categorized in 4 main Categories;

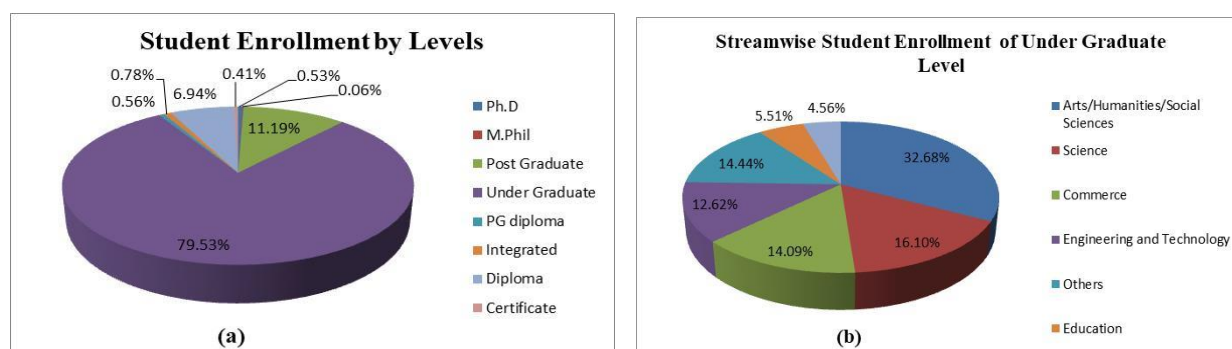
University, College, Schools, and Stand-Alone Institutions. Table 1 shows the statistics of total number of universities, colleges, and standalone Institutions in India and Maharashtra state.

**Table 1:** Total number of universities, colleges and standalone Institutions in India and Maharashtra state

	Universities	Colleges	Stand Alone Institutions
India	1043	42343	11779
Maharashtra	46	4494	6952

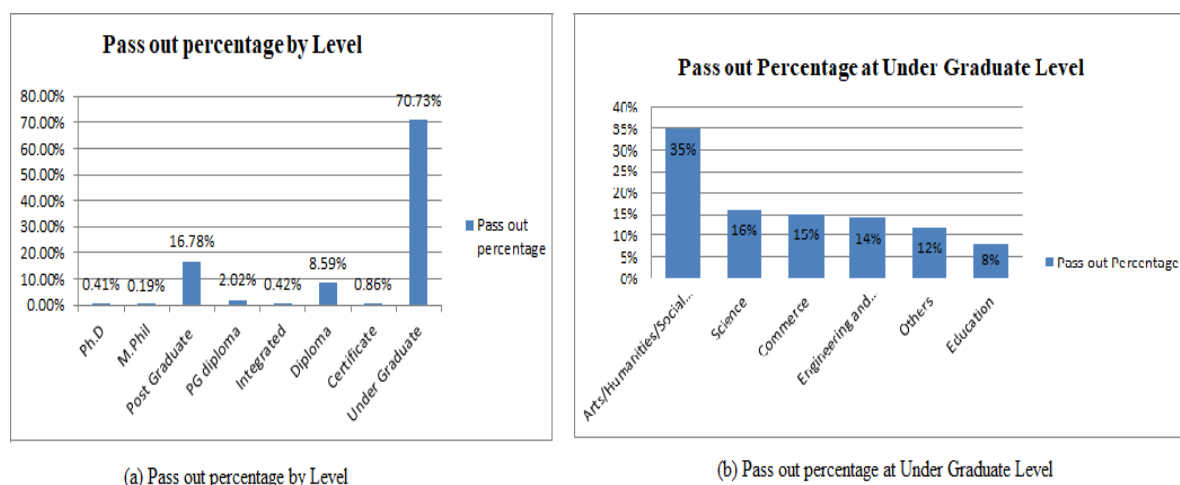
(Source: AISHE Report 2019-20)

Gross Enrollment Ratio (GER) is the total enrolment in a specific level of education, regardless of age, and expel. It is a percentage of the population in the official age group corresponding to the level of education (Joshi & Ahir, 2019). In India, the GER in higher education is 27.1, calculated for the 18-23 age groups. Total enrollment in higher education is estimated to 38.5 million (Ministry of Education, Government of India, AISHE Report 2019-20)(Figure 1).



**Figure 1:** (a) Student Enrollment by Levels (b) Student Enrollment of Under Graduate Levels (Source: AISHE Report 2019-20)

Percentage share of student Out-Turn in different levels as well at undergraduate level are shown in Figure 2.



**Figure 2:** Student pass out percentage (a) by level (b) at Under Graduate Level (Source: AISHE Report 2019-20)

As per the above statistics 79.53% student enrolled for undergraduate level and student passing percentage is 70.73% (Ministry of Education, Government of India, AISHE Report 2019-20). There are 30,619,050 students enrolled for undergraduate level programs, and 27,231,050 students completed graduation successfully (Ministry of Education, Government of India, AISHE Report 2019-20). It is clear that students enrolled for undergraduate courses but approximately 10% of student failed to complete the graduation. Student academic success ratio in science and engineering and technology streams is comparatively low as compared to arts stream (Ministry of Education, Government of India, AISHE Report 2019-20). There is need to identify the causes of deterioration.

### a) Role of Machine Learning for SASP

In their study, the authors (Martínez-Carrascal et al., 2020) tries to find out the impact of early activity on flipped classroom performance. Different algorithms are used for prediction such as Decision Tree (DT), KNN, SVM. The precision method is used to check the accuracy of algorithms' prediction, and SVM gives the highest accuracy. Indicators such as class id, item id, action expected, due date, quiz on time, quiz total, lectures on time, lecture total, log lines, login days, attendance, num\_mat were used for the study. The study concluded that model works good for fresher as compared to repeater student as well as it is found that student engagement in class either attendance or quiz completion are critical factors in success. The study is suitable for predicting success and with large dataset and with additional indicators better prediction can be made.

Authors (Akash et al., 2021) tried to find the correlation between co-curricular activities and student academic performance. The techniques used were voting perceptron, logistic regression, MLP, and the random forest classifier. The study stated that logistic regression gives highest accuracy and with correlation of 69% between CGPA and extracurricular activities. The extracurricular activities are major factor in student success prediction and need to be explored more with proper and in depth study.

The study by authors (Aviso et al., 2021) focuses on finding the influential institutional indicators which affect graduate employability using classification models. The indicators used for study were international research network, inbound and outbound exchange, papers and citations per papers, international students, international faculty, highly qualified faculty, teacher-student ratio, academic reputation, employer reputation as a decision variable. The study derived and validates five rule-based classifiers, and it is concluded that there is a significant alliance between research and internalization with employability. The future scope is high for this research, and additional indicators might result in a different association.

### b) Role of Machine Learning for SASP in Indian Context

The review by (Kumar et al., 2017) stated that internal marks and CGPA of the student are important indicator for the prediction of results. Also, in the review author concluded that most of the researchers used classifiers such as DT, NB, and Rule-Based algorithms for predicting student's academic performance.

Author (Dhilipan et al., 2021) proposed a prediction system for student identification. Binomial logical regression, DT, and KNN classifier machine learning techniques used for the study. Performance of each algorithm was evaluated using measuring metrics such as confusion matrix, precision, recall, f1-score support. The indicators used for study were 10<sup>th</sup> marks, 12<sup>th</sup> marks and three semester marks. Binomial logical regression gives highest accuracy 97.05%. Additional features can be used for better prediction.

In one of the study authors (Pallathadka et al., 2021) used Nave Bayes, ID3, C4.5, and SVM algorithms to classify and predict student performance. These algorithms are analyzed with accuracy and error rate metrics and it is found that SVM gives highest accuracy 88% for prediction. UCI machinery student performance data set (Cortez & Silva, 2008) was used in the study. Table 3 gives brief review of machine learning algorithm used for the SASP in Indian Context.

## 7. Conclusion

At the end of the research paper, a well-integrated framework of the influential indicators will be provided. Using different machine learning techniques, a well-defined prediction model will be designed for a student's early success prediction in higher education. Interactive decision support will be designed, which will help different stakeholders to make timely decisions to solve different problems.

## References

1. A., V., D., P., & V., M. (2017). Predicting Student's Performance using Machine Learning. *Communications on Applied Electronics*, 7(11), 11–15. <https://doi.org/10.5120/cae2017652730>
2. Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2), e01250. <https://doi.org/10.1016/j.heliyon.2019.e01250>
3. Akash, P. P., Parvin, M., Moon, N. N., & Nur, F. N. (2021). Effect of Co-curricular activities on student's academic performance by machine learning Authors: Shaikh Rezwan Rahman 1 Department of Computer Science & Engineering, Daffodil International University. *Current Research in Behavioral Sciences*, 100057. <https://doi.org/10.1016/j.crbeha.2021.100057>
4. Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review Available Online At: <https://jazindia.com>

- and best practices. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-0177-7>
5. Anuradha, C., & Velmurugan, T. (2015). A comparative analysis on the evaluation of classification algorithms in the prediction of students performance. *Indian Journal of Science and Technology*, 8(15). <https://doi.org/10.17485/ijst/2015/v8i15/74555>
  6. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
  7. Aviso, K. B., Iii, F. P. A. D., Janairo, J. I. B., Lucas, R. I. G., Promentilla, M. A. B., Tan, R. R., & Yu, D. E. C. (2021). *What university attributes predict for graduate employability?* 2(February), 1–8. <https://doi.org/10.1016/j.clet.2021.100069>
  8. Borah, D., Malik, K., & Massini, S. (2021). *Teaching-focused university – industry collaborations : Determinants and impact on graduates ' employability competencies*. 50(March 2020). <https://doi.org/10.1016/j.respol.2020.104172>
  9. Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *15th European Concurrent Engineering Conference 2008, ECEC 2008 - 5th Future Business Technology Conference, FUBUTEC 2008, 2003(2000)*, 5–12.
  10. Dhilipan, J., Vijayalakshmi, N., Suriya, S., & Christopher, A. (2021). Prediction of Students Performance using Machine learning. *IOP Conference Series: Materials Science and Engineering*, 1055(1), 012122. <https://doi.org/10.1088/1757-899x/1055/1/012122>
  11. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, 25, 326–332. <https://doi.org/10.1016/j.protcy.2016.08.114>
  12. Joshi, K. M., & Ahir, K. V. (2019). Higher education in India: Issues related to access, equity, efficiency, quality and internationalization. *Academia (Greece)*, 2019(14), 71–91. <https://doi.org/10.26220/aca.2979>
  13. Khan, I., Al Sadiri, A., Ahmad, A. R., & Jabeur, N. (2019). Tracking student performance in introductory programming by means of machine learning. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS C 2019*, 1–6. <https://doi.org/10.1109/ICBDSC.2019.8645608>
  14. Kishan Das Menon, H., & Janardhan, V. (2020). Machine learning approaches in education. *Materials Today: Proceedings*, 43(xxxx), 3470–3480. <https://doi.org/10.1016/j.matpr.2020.09.566>
  15. Kovacic, Z. J. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. *Proceedings of the 2010 InSITE Conference*, 647–665. <https://doi.org/10.28945/1281>
  16. Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., Gonyea, R. M., & Gonyea, M. (2012). *Unmasking the Effects of Student on First-Year College Engagement Grades and Persistence*. 79(5), 540–563.
  17. Kumar, M., Singh, A. J., & Handa, D. (2017). Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques. *International Journal of Education and Management Engineering*, 7(6), 40–49. <https://doi.org/10.5815/ijeme.2017.06.05>
  18. Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1(9), 1–10. <https://doi.org/10.1007/s42452-019-0884-7>
  19. Mardis, M. A., Ma, J., Jones, F. R., Ambavarapu, C. R., Kelleher, H. M., Spears, L. I., & McClure, C. R. (2018). Assessing alignment between information technology educational opportunities, professional requirements, and industry demands. *Education and Information Technologies*, 23(4), 1547–1584. <https://doi.org/10.1007/s10639-017-9678-y>
  20. Martínez-Carrascal, J. A., Márquez Cebrián, D., Sancho-Vinuesa, T., & Valderrama, E. (2020). Impact of early activity on flipped classroom performance prediction: A case study for a first-year Engineering course. *Computer Applications in Engineering Education*, 28(3), 590–605. <https://doi.org/10.1002/cae.22229>
  21. Mesarić, J., & Šebalj, D. (2016). *Decision trees for predicting the academic success of students*. 7, 367–388. <https://doi.org/10.17535/corr.2016.0025>
  22. Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51. <https://doi.org/https://doi.org/10.1016/j.dss.2018.09.001>
  23. Nhamo, G., & Mjimba, V. (n.d.). *Sustainable Development Goals and Institutions of Higher Education*.
  24. Pallathadka, H., Wenda, A., Ramirez-asís, E., Asís-lópez, M., Flores-albornoz, J., & Phasinam, K. (2021). *Materials Today : Proceedings Classification and prediction of student performance data using*

- various machine learning algorithms. *Materials Today: Proceedings*, xxxx. <https://doi.org/10.1016/j.matpr.2021.07.382>
25. Rathee, A., Mining, D., Mining, E. D., & Algorithm, C. (2013). Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance ID3 Algorithm. *International Journal of Computers & Technology*, 4(2), 244–247.
  26. Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2(March), 100018. <https://doi.org/10.1016/j.caeai.2021.100018>
  27. Roksa, J., & Kinsley, P. (2018). The Role of Family Support in Facilitating Academic Success of Low - Income Students. *Research in Higher Education*, 0123456789. <https://doi.org/10.1007/s11162-018-9517-z>
  28. Salah Hashim, A., Akeel Awadh, W., & Khalaf Hamoud, A. (2020). Student Performance Prediction Model based on Supervised Machine Learning Algorithms. *IOP Conference Series: Materials Science and Engineering*, 928, 032019. <https://doi.org/10.1088/1757-899x/928/3/032019>
  29. Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. *ACM International Conference Proceeding Series, Part F1481*, 7–11. <https://doi.org/10.1145/3318396.3318419>
  30. Sripath Roy, K., Roopkanth, K., Uday Teja, V., Bhavana, V., & Priyanka, J. (2018). Student career prediction using advanced machine learning techniques. *International Journal of Engineering and Technology(UAE)*, 7(2), 26–29. <https://doi.org/10.14419/ijet.v7i2.20.11738>
  31. Suresh, A., S, B. S., R, E. K., & N, G. (2020). Student Performance Prediction using Machine Learning. *International Journal of Computer Science and Mobile Computing*, 9(9), 38–42. <https://doi.org/10.47760/ijcsmc.2020.v09i09.004>
  32. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences (Switzerland)*, 10(6). <https://doi.org/10.3390/app10062145>
  33. Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2020). The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology*, 51(2), 347–370. <https://doi.org/10.1111/bjet.12836>
  34. Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs. *IEEE Journal on Selected Topics in Signal Processing*, 11(5), 742–753. <https://doi.org/10.1109/JSTSP.2017.2692560>
  35. Yang, H., Ph, D., Su, J., Ph, D., Bradley, K. D., & Ph, D. (2020). *Applying the Rasch Model to Evaluate the Self-Directed Online Learning Scale ( SDOLS ) for Graduate Students*. 21(3).
  36. Yıldız Aybek, H. S., & Okur, M. R. (2018). Predicting Achievement with Artificial Neural Networks: The Case of Anadolu University Open Education System. *International Journal of Assessment Tools in Education*, 5(3), 474–490. <https://doi.org/10.21449/ijate.435507>
  37. York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research and Evaluation*, 20(5), 1–20.
  38. Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers and Electrical Engineering*, 89(November 2020), 1–10. <https://doi.org/10.1016/j.compeleceng.2020.106903>