Machine Learning-based Model to Predict Student's success in Higher Education

Atul Garg[1], Nidhi Bansal Garg[2], Umesh Kumar Lilhore[3], Renu Popli[4], Ankit Bansal[5]

{atul.garg@chitkara.edu.in[1], nidhphy@gmail.com[2], umeshlilhore@gmail.com[3]}

Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India[1,2,4,5]
Department of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali, Punjab, India[3]

Abstract. Predictions are always helpful for making decisions. Students are the future of the world. Higher Education Institutions (HEI's) in developing countries cannot apply similar strategies to all the students. Academic achievement plays a crucial role in the academic system because it is often utilized for the educational establishment quality. Early identification of at-risk educators and prevention strategies can significantly improve their chances of succeeding. Education is affected by different environments, family backgrounds, social and personal responsibilities. In this research article student's performance is measured based on various parameters using Random Forest, Naive Bayes and K* method. Experimental analysis shows the strengthening of the random forest method over K* and Naive Bayes method.

Keywords: Naive Bayes, K*, Students performance, HEI's, machine learning, Big Data.

1 Introduction

Human has been learning since the time of nature's evolution. After, the development of civilization, many developments happened. All these developments could be possible with the learning methods. If the education system is compared with the ancient time, anyone can see the development in the human being. Even with the revolution of Internet and Web, every area of society, education, business, agriculture is affected.

After the success of the Internet and Web, the revolution of smartphones changed everything. Before smartphones, only professionals and educated people were able to use the Internet. However, with smartphones, everyone, whether he/she is child/young/aged/senior citizen or educated/illiterate are using the Internet. Information, communication is just far from a single fingure touch. Even the cities, transport, shopping, kitchen are also becoming smarter [1],[2].

This revolution has also affected the education system. The learning pattern for the students and the teaching methodologies of teachers are also upgraded with the revolution of technology. The Student's success is depend on various aspects, e.g., geographical area, family background, poverty, education of parents, level of school, gender, mode of education.

In India, the aspects mentioned earlier plays a vital role in education. India is a big country that is divided into different geographical areas. Furthermore, the schooling facilities are also different. Family background in the overall growth of student also matters. Money is another essential requirement in studies. It can be seen in developing countries like Inida, that the poor parents are unable to pay school fees and the children have to leave schooling to earn. The educated parents can guide and help their children in their studies.

Whereas, illiterate parents can't guide properly about education to their children, it is difficult to guide them for further studies. Moreover, in the absence of proper guidance, the children cannot get proper directions. Further, the level of school and teachers also affect the growth of the students. The school environment, well-educated teachers, proper laboratories, curriculum and extra-curricular are too helpful for the overall growth of the students. However, in its absence, the overall growth of students may affect.

In India, the gender of the children also matters in their growth. Many parents do not want to educate their daughters either their daughters have good calibre after schooling or they have enough money to educate. After COVID-19, the model of education is changed. Teaching and learning methodology converted from physical classes to online classes. Developing countries like India do not have enough resources for online classes.

Government is also providing various facilities to the people related to education system to overcome all the aspects e.g., no fees in Government school, mid-day-meal program, beti-bachao beti-padhao, education loan etc. But, these all points directly or indirectly affect the growth of students.

The important reason of unsuccessful students is lack of prediction about students. Generally, the schools, HEI's doesn't bother to know about the background of the students. Rather, if the Student's information is gathered earlier then the HEI's can plan the proper strategies for the overall growth of the students. Big Data Analytics, AI and Machine Learning techniques are playing a vital role to predict the future of the students. Which can help the HEI's to take further decisions and actions for needy students. In this work student's performance is analyzed on the basis of their placements.

2 Literature Review

Since the development of human the learning process is going on. Learning is endless and continuous research is going on to make it more efficient. Predictions are playing good role in the development. In this section research work done by numerous researchers is discussed.

The authors [3] focus on profiling of the students and also focused on both online and off line education. They focused on their behaviours during their online and offline classes. According to authors [4] the main characteristics of students are behaviour, unique identity and academic records. Two data sets are used on different classifications to verify their results. And, the authors found the best results with decision tree.

The authors [4] focused on the overall success of students. According to authors [4], for the success of students various higher education Institutions apply different strategies. And, to apply these strategies a good prediction is most important. And, with the development of Big Data, AI and ML prediction became easy. In their work the authors [4] proposed automated machine learning technique to predict the Student's performance.

According to authors in [5],[6], the academic Institution's success depends on the success of the students. If the students are not successful then the Institution's admission may decrease.

By authors [3],[7], the students could be more successful if their drawbacks, risks problems can be detected earlier. Once the different risks are identified their solutions could be found on time

and ML techniques are useful to detect early risks. In their study, the author [5] insist on Education Data Mining (EDM) [1],[8]technique for prediction.

Authors in[6] focused on small datasets. Key Indicators were used for small datasets to create prediction model. The authors [6] used datasets on various ML algorithms and they found better results using support vector machine and learning discriminant analysis algorithms.

Authors in[9],[10], [11] focused on artificial and ML technologies. According to them, these techniques have changed the world with the help of predictions. Also, these techniques made the education system more transparent, intelligent and more efficient. Different classification methods were applied on different datasets to predict to success of the students.

The authors in [12], [13], [14], [15] focused on internal assessment and Cumulative Grade Points Average (CGPA) to predict the performance of students. The authors [12], [13] presented that the grad points are more important to predict Student's future regarding placement, area of interest and future opportunities. In their study the authors [13]applied dataset on neural network and decision trees to prove their study. Age, disability, family background, gender geographical location and other achievements of students are considered as main attributes by researchers[15], [16], [17]in their study. The selected attributes plays an important role in teaching and learning process of one's life.

The authors in [16], [17], [18], [19] focused on the gender of the students. According to [16], [17], [18], [19] male students are not more education centric and hardworking than female. The authors in [20], [21] also selected other attributes of one's life e.g. behaviour in school, social responsibilities and life, participation in other activities e.g. sports, cultural, competitions etc.

The authors [22] categorized the students in slow, medium and advance learners. According to authors [22], different strategy need to apply on different kind of students. Some special classes, motivational talks and counselling should be given to slow learners, counselling and motivation should be applied on medium learners whereas some special talks on advance technology and motivation should be applied with some appreciation. So that other students could set their goal. Further, classification techniques of ML were applied on their dataset to show the results. In Table 1 comparative analysis of key methods, softwares and the parameters used by renowned researchers are discussed.

Table 1: Comparative analysis of various existing works

Reference	Key Methods	Key parameters, Results	Software
[1]	J 48, SCM	Precision, Accuracy, Recall, F-measure	WEKA
[2]	ID3, KNN	Precision, Accuracy, Recall	Rapid-Miner
[3]	PNN, NB	Precision, Accuracy, Recall	Rapid-Miner
[4]	LR, NB	Precision, Accuracy, Recall	WEKA
[5]	SVM, J-48, NB	Precision, Accuracy, Recall	Python Anaconda
[6]	RF, J-48	Precision, Accuracy, Recall	WEKA
[7]	NB, SVM	Precision, Accuracy, Recall	Rapid-Miner
[8]	Ensemble Methods	Precision, Accuracy, Recall, F-measure	WEKA
[9]	J-48, K*	Precision, Accuracy, Recall	Python Anaconda

3 Methodologies

In the previous section, it can be seen that various researchers discussed about the success of the students. Different researchers focused on different attributes to predict the future of the students. The main aim of study is to recognise the weak areas of the students before time so that a HEI's can apply their strategic plan on time. It will be also helpful to identify the dropout students.

Datasets are collected from Kaggle [23], [24], [25] for this study. Information of 215 students with 15 attributes (Figure 1) are consisted in this dataset. The values of dataset e.g., no of instances, minimum, maximum, mean and standard deviation values are depicted in Figure 2. This dataset is applied on Waikato Environment for Knowledge Analysis (WEKA) version 3.8.5 for results.
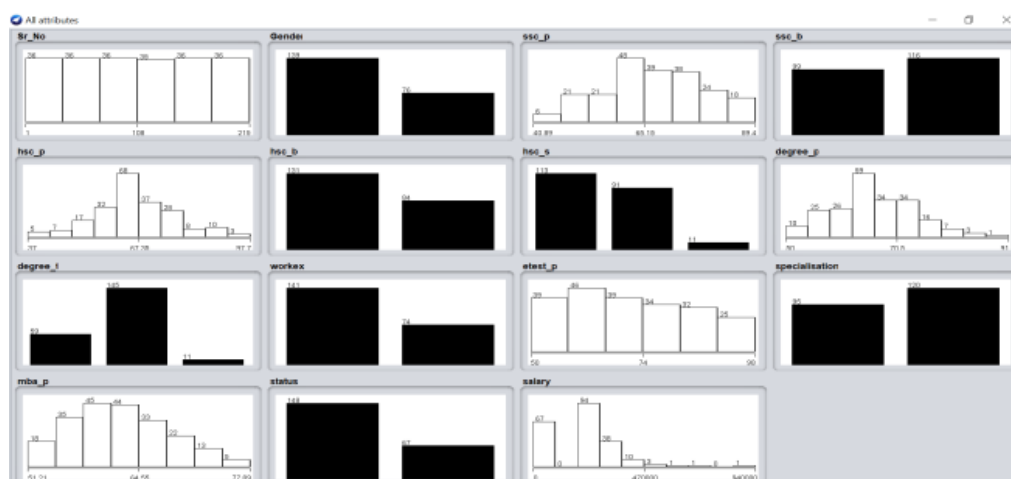

Fig. 1. All attributes in dataset

The above discussed dataset is applied on the following classifiers.

A. Naive Bayes Naïve Bayes is powerful and simple algorithm used to prediction which is appropriate for multiclass and binary classification. In this classification algorithm class selection is done on the basis of probability. Considered class should have highest probability. This classifier model assumes each variable as an independent variable. Naïve Bayes is very effective on complex and large range of problems. The above discussed dataset is applied on Naïve Bayes.

B. K* It is a heuristic Search Algorithm and used to find k shortest paths. K* algorithm is improved from its conventional algorithms. As the K* can be guided with heuristic algorithms and works on the fly so it provides better results than the conventional algorithm.

C. Random forest A random forest (RF) is a type of machine learning which's utilized to overcome classification and regression challenges. It implements ensemble techniques that

are a technology that integrates numerous classification models to deliver alternatives to complicated complications. A random forest classifier contains of several decision trees.
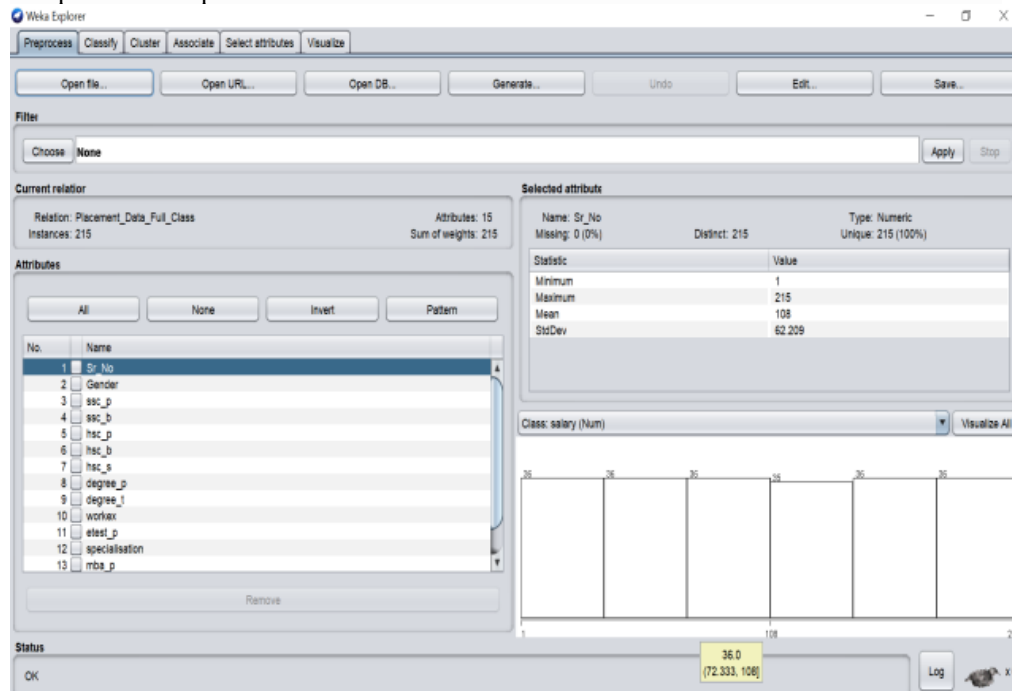

Fig. 2. Dataset values

4 Results and Discussion

This investigation discussed the possibility of forecasting university students' performance the premise of various input factors extracted in Higher education institutions. The methodology was built utilizing supervised machine learning techniques (Naive Bayes, Random Forest and K*, and also the comparison is made. The results are discussed in Table 2 and depicted in Figure 3.

Table 2: Experimental Results comparison

Method	Precision	Recall	TP	FP
Random Forest	0.891	0.921	0.787	0.139
Naive Bayes	0.812	0.845	0.767	0.259
K*	0.787	0.814	0.731	0.524

An experimental result shows Precision value 0.891, Recall 0.921, TP 0.787 FP 0.139 for Random forest. Naive Bayes shows, Precision value 0.812, Recall 0.845, TP 0.767 and FP 0.259

and K* method shows Precision 0.787, Recall 0.814, TP 0.731 and FP 0.524. An experimental analysis clearly indicates that Random Forest classifiers performs outstanding in terms of Precision, Recall, TP and FP as compared to Naive Bayes and K* Classifiers.
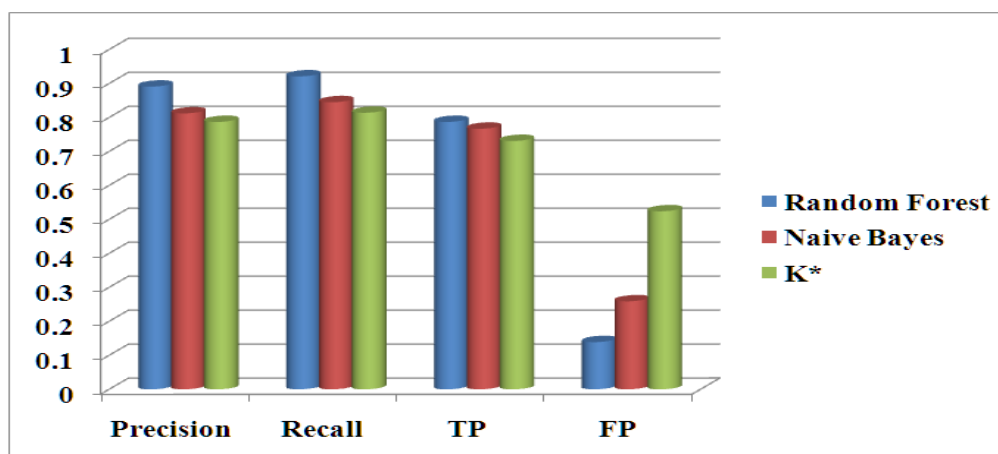

Fig. 3. Experimental Results comparison

5 Conclusions

Students are the future of the country and society The success of the student's directly or indirectly impact on the coutry, society and HEI's. The successful students can improve the economy of country and different rankings of the HEI's also depend on the successful students. In this study, the different parameters for overall growth of the students are discussed. And, the similar dataset is applied on three methods named as NaiveBayes, Random Forest and K*. The better results are achieved with random forest method. It can be conclude that the early prediction of the students success can provide better results.

References

[1] Dutt A., Ismail M. A., and Herawan T.: A Systematic Review on Educational Data Mining. IEEE Access. Vol. 5, No. c, pp. 15991–16005 (2017). doi: 10.1109/ACCESS.2017.2654247.

[2] Chhabra R., Verma S. and Krishna R.: A survey on driver behaviour detection techniques for intelligent transportation systems. 7th Int. Conf. Cloud Comput. Data Sci. Eng., Vol. 7, IEEE-2017, pp. 36–41 (2017).

[3] Hamim T., Benabbou F. and Sael N.: Survey of Machine Learning Techniques for Student Profile Modelling. Int. J. Emerg. Technol. Learn., Vol. 16, No. 4, pp. 136–151 (2021).
doi: 10.3991/ijet.v16i04.18643.

[4] Zeineddine H., Braendle U., and Farah A.: Enhancing prediction of student success: Automated machine learning approach. Comput. Electr. Eng., Vol. 89, No. April 2020, pp. 106903 (2021).
doi: 10.1016/j.compeleceng.2020.106903.

[5] Alyahyan, E., Düştegör, D. Predicting academic success in higher education: literature review and best practices. Int J Educ Technol High Educ Vol 17, No 3, pp. 1-21 (2020). https://doi.org/10.1186/s41239-020-0177-7

[6] Zohair L. M. Abu: Prediction of Student's performance by modelling small dataset size. Int. J. Educ. Technol. High. Educ., Vol. 16, No. 1, pp. 1-18 (2019). doi: 10.1186/s41239-019-0160-3.

[7] Garg, A., Popli, R., & Sarao, B. S.: Growth of digitization and its impact on big data analytics. In IOP Conference Series: Mat. Sci. and Engg., vol. 1022, no. 1, pp. 012083, 2021.

[8] Liñán L. C. and Pérez Á. A. J.: Minería de datos educativos y análisis de datos sobre aprendizaje: Diferencias, parecidos y evolución en el tiempo. RUSC Univ. Knowl. Soc. J., Vol. 12, No. 3, pp. 98–112 (2015). doi: 10.7238/rusc.v12i3.2515.

[9] Garg A. and Amin D. M.,: A Proposed Framework Using Neural Network in Web Mining for Improving the Performance of E-Learning System. Int. J. Sci. Res. Vol. 12, No. 12, pp. 2681–2688, (2014).

[10] Halde R. R.: Application of Machine Learning algorithms for betterment in education system. Int. Conf. Autom. Control Dyn. Optim. Tech. ICACDOT 2016, pp. 1110–1114 (2017).
doi: 10.1109/ICACDOT.2016.7877759.

[11] Tsai S. C., Chen C. H., Shiao Y. T., Ciou J. S., and Wu T. N.: Precision education with statistical learning and deep learning: a case study in Taiwan. Int. J. Educ. Technol. High. Educ. Vol. 17, No. 1, pp. 1-13 (2020). doi: 10.1186/s41239-020-00186-2.

[12] M. Quadri and D. Kalyankar: Drop out feature of student data for academic performance using decision tree techniques. Glob. J. Comput., Vol. 10, No. 2, pp. 2–5 (2010).

[13] Shahiri A. M., Husain W., and Rashid N. A.: A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Comput. Sci., Vol. 72, pp. 414–422 (2015).
doi: 10.1016/j.procs.2015.12.157.

[14] Ashraf A., Anwer S. and Khan M. G. A Comparative Study of Predicting Student' s Performance by use of Data A Comparative Study of Predicting Student' s Performance by use of Data Mining Techniques. Am. Sci. Res. J. Eng. Technol. Sci., Vol. 44 No.1, pp. 122–136 (2018).

[15] Ibrahim Z. and Rusli D.: Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision tree and Linear Regression. Proc. 21st Annu. SAS Malaysia Forum, no. September, pp. 1–6, (2007).

[16] Bin Mat U., Buniyamin N., Arsad P. M., and Kassim R. A.: An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. 2013 IEEE 5th Int. Conf. Eng. Educ. Aligning Eng. Educ. with Ind. Needs Nation Dev. ICEED 2013, pp. 126–130 (2014). doi: 10.1109/ICEED.2013.6908316.

[17] Christian T. M. and Ayub M.: Exploration of classification using NBTree for predicting students' performance. Proc. 2014 Int. Conf. Data Softw. Eng. ICODSE 2014, pp. 0–5, (2014). doi: 10.1109/ICODSE.2014.7062654.

[18] Gupta, M. and Garg, A., 2015: A Comparative Analysis of Content Delivery Network and Other Techniques for Web Content Delivery. International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), Vol. 6, No. 4, pp.43-58 (2019).

[19] Simsek A. and Balaban J.: Learning Strategies of Successful and Unsuccessful University Students. Contemp. Educ. Technol. Vol. 1, No. 1, pp. 36–45 (2020)., doi: 10.30935/cedtech/5960.

[20] Amin, D.M., Garg, A.: Performance analysis of data mining algorithms. J. Comput. Theor. Nanosci. Vol. 16, No. 9, 3849–3853 (2019).

[21] Hooshyar D., Pedaste M. and Yang Y.: Mining educational data to predict students' performance through procrastination behavior. Entropy, Vol. 22, No. 1, p. 12 (2020). doi: 10.3390/e22010012.

[22] Garg A., Lilhore U. K., Ghosh P., Prasad D. and Simaiya S.: Machine Learning-based Model for Prediction of Student's Performance in Higher Education. 8th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 162-168 (2021).

doi: 10.1109/SPIN52536.2021.9565999.

[23] Roshan B.: Academic and Employability Factors influencing placement. [Online] Available: https://www.kaggle.com/benroshan/factors-affecting-campus-placement.

[24] Katarya R., Gaba J., Garg A. and Verma V.: A review on machine learning based student's academic performance prediction systems. International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE Xplore, pp. 254-259 (2021).

[25] Gonzalez-Nucamendi A., Noguez J., Neri L., Robledo-Rella V., García-Castelán R. M. G. and Escobar-Castillejos D.: The prediction of academic performance using engineering student's profiles. Comput. Electr. Eng., Vol. 93, No. 107288, p. 107288 (2021).