

Predicting Student Performance Using Machine Learning Approach

Ankita B Bhor¹, Rituja V Dhamdhare², Nikita A Wable³, A.S. Kharat⁴

^{1,2,3}B.E Students, ⁴Professor,

^{1,2,3,4}Department of Computer Engineering, Jaihind College Of Engineering, Kuran, Pune, India

Abstract- Predicting Students performance beforehand can be very beneficial for educational institutions to improve their teaching quality[3]. When instructors do receive inadequate feedback concerning a course session or week, the result can be that students fail the course, drop out, or get a lower grade on the final exam. This paper proposes to predict students performance by considering their academic details. We use the classification algorithm to predict the performance of student. Educational organizations are unique and play important significant role for the development of any country. As Education transforms the lives of individuals, families, communities, societies, countries and ultimately the world![5]. This is why we live comfortable lives today. Now a day's education is not limited to only the classroom teaching but it goes beyond that like Online education System, Web-based Education System, Seminars, Workshops, MOOC course[6]. becomes It's more challenging to Predict student's performance because of the huge bulks of data stored in the environments of Educational databases, Learning management databases[4]. Students' performance can be evaluated with the help of various available techniques[1]. It is evolving area of study that emphasizes on various techniques like classification, prediction, feature selection[11]. In addition, the result show that employed on learning records or data related to education to predict the students' performance and learning behavior by extracting the hidden knowledge.

Keywords- Performance Prediction , Machine Learning, Classification, Predictive Model, Database, Classification algorithms.

I. INTRODUCTION

Today every educational institution handles and deals with large amount of student data which can be beneficial for a number of reasons. One of the important application of such data is predicting student performance. Such a prediction can be useful not only for the students but also for teachers/mentors. Mentors can provide special assistance to the students who are on the verge of failing. In order to determine which category a student lies, such data can be quite helpful. This application can be used by any prominent school or colleges. It can be used to predict the pointer ranges or percentage range for future semester exams. These ranges can be predicted using a number of data mining algorithms such as classification algorithms, rule-based algorithms,

ensemble methods, and neural networks. The main aim of this project is the selection of features that show a strong relationship with a target attribute that is to be predicted from a high dimensional data-set.

We have evaluated and compared the number of algorithms such as decision tree, random forest, support vector machine, naive Bayes and neural networks by applying them on the data-set. The rest of the paper provides an explanation on nature of neural networks along with the results of our evaluation. Machine learning is used for analyzing data based on past experience and predicting future performance .Reinforcement machine learning algorithms is a branch of artificial intelligence .It automatically determines the behavior of environment and maximizes its performance .In this algorithm agent is used for taking decision . With the help of agent best action is selected .It gives feedback by adapting the environment. It is an automated learning system .In this system we give input in the terms of marks and attendance. The intention of our system is to analyze and predict performance of student that will help to improve the marks of that student. In this paper this system is implemented for helping student to improve their academic performance.

II. LITERATURE REVIEW

It has used DT and BN classes of MLAs for predicting the undergraduate and post graduate results of two universities in Thailand. The student records used for this prediction is 20492 and 932 respectively. This algorithm used for predict C4.5, MSP and Naïve Bayes. The conclusion for all classes of predictions DT yields better results than BN by 3 to 12%. Resampling was used to improve the prediction accuracy[1].

It described a model to predict student results for a distance learning course in Hellenic Open University. Predictions were done on the basis of marks obtained in written assignments. The algorithms used for this prediction are C4.5, Naïve Bayesian Network (NBN), Back Propagation (BP), 3-Nearest Neighborhood (3-NN) and Sequential Minimal Optimization (SMO). A set of 510 students of the university was chosen for experimental purpose. It was found that the NBN algorithm generates the best results (accuracy 72.48%)[4].

In that developed a predictive data mining model for student performance to identify the factors causing poor performance

in higher secondary examination in Tamil Nadu. A data set for 772 students collected from regular students and school offices were used for this prediction. Algorithm used for this prediction is Chi-Square Automatic Interaction Detection (CHAID) DT. This tree was used to generate a set of decision rules used for predicting student grades. The overall prediction accuracy achieved was 44.69%[5].

They applied data mining algorithms on “logged data” in a educational web based learning system. The system was tested with a data set of 227 students enrolled in a physics course in Michigan State University. Classification was initially performed using Quadratic BN, 1-NN, Prazen Window, Multilayer Perceptron (MLP) and C5.0 DT. It was seen that combining these classifiers increases prediction accuracy. Genetic Algorithms (GA) were further used to improve prediction accuracy by 10%[6].

It explores the “socio-demographic” and “study environment” factors that results in student dropout in a polytechnic college in New Zealand. He uses student enrollment data like age, gender, ethnicity for this purpose. The total number of student records used for purpose was 450. Algorithms used for this prediction are CHAID and Classification and Regression Trees (CART). It was found that CART obtained a higher degree of accuracy (60.5%). Based on the results of Confusion Matrix and ROC curve he concluded that decision trees based on enrollment data alone are not sufficient to classify students accurately[7].

They are used ANN for predicting student graduation outcomes at Waubensee College. The prediction model was constructed from a profile of 1407 students of which 1100 were used for training and 307 were used for testing purpose. The average predictive efficiency for training and test sets were 77% and 68% respectively[8].

In current system we use the academic data set of students and then apply the machine learning techniques. i.e random Forest ,Naïve bayes to predict students performance.

III. EXISTING SYSTEM

A machine learning Approach for tracking and predicting student performance in degree programs in this paper we were only tracking the degree programs and not predicting the overall performance of students and also they were using a data driven approach and clustering method based on probabilistic matrix factorization this is quite difficult to predict the student performance and ensemble learning algorithms were used which is hard to implement And complex data set is allowed.

IV. PROPOSED SYSTEM

The proposed system automatically handles student’s data to predict their performance. System predict the defaulter and non-performing students.

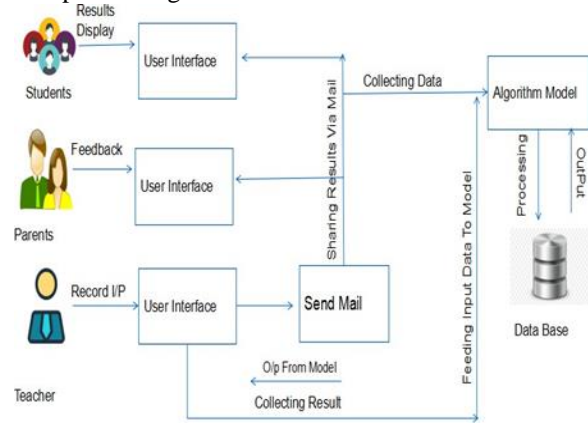


Fig.1: System Architecture

A. Student:

Students can only view their performance analysis and can focus on to suggested points.

B. Teacher:

The teacher will be giving students academics curriculum as an input to the system as well as can fill up the students details and parents details, can generate results and share it to the concerned students’ parents.

C. Parent:

The parents can log in and can view student’s data analysis and can share the feedback’s regarding the analyzed reports.

V. CONCLUSION

Predicting student performance based on their on-going academic records. Also we conclude that proposed system is helping us to make the student performance better. In this paper machine learning can prove to be powerful tool and all algorithms we used increases of data set size.

VI. REFERENCES

- [1]. N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. SchmidtThieme, “Recommender system for predicting student performance,” *Procedia Com-puter Science*, vol. 1, no. 2, pp. 2811–2819, 2010.
- [2]. Castro, F., Vellido, T., Àngela Nebot, and Mugica F., *Applying Data Mining Techniques to e-Learning Problems*.
- [3]. Nguyen, N. N., Janeck, P, Haddawy, P., 2007, A Comparative Analysis of Techniques for Predicting Academic Performance, 37th ASEE/IEEE Frontiers in Education Conference.
- [4]. Kotsiantis, S., Piarrekeas, C., Pintelas, P.,2007.Predicting Students’ performance in Distance Learning using Machine Learning Techniques, *Applied Artificial Intelligence*, 18:411-426.

- [5]. Ramaswami, M., Bhaskaran, R.,2010, A CHAID Based Performance Prediction Model in Educational Data Mining, International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1.
- [6]. Minei-Bidgoli, B., Kashy, D., Kortemeyer G., Punch W F, 2003. Predicting Student Performance: An Application of Data Mining Methods with the Educational Web-Based System LON-CAPA, 33rd ASEE/IEEE Frontiers in Education Conference,.
- [7]. Kovacic, Z. J.,2007, Early Prediction of Student Success: Mining Students Enrolment Data, Informing Science &IT Education Conference .
- [8]. Karamohzis and Vrettos online and no one sees it, is it still an ad?" Journal of Advertising Research, 2013.
- [9]. Livieris, E., Drakopoulou, E., Pintelas, P., Predicting students' performance using artificial neural networks.
- [10].Stefanowski, J., An Experimental Study of Methods Combining Multiple Classifiers - Diversified both by Feature Selection and Bootstrap Sampling.
- [11].Zhao, Y., and Zhang, Y.,Comparison of decision tree methods for finding active objects.
- [12].Acharya, A., Sinha, D., 2014, Application of Feature Selection Methods in Educational Data Mining, International Journal of Computer Applications