Machine Learning Based Student Performance Analysis

Dr Himanshu Maniar

BCA Department Shri C U Shah College of Commerce, Management and Computer Education

Wadhwan-363030 Gujarat India

ABSTRACT:

Over the years, the performance analysis of students has become an important activity for the academicians to identify the quality of education. The same has been carry out by the government and recruiters for their own analysis. The traditional methods of analysis are still in practice but have found to be less effective for accurate information. Machine Learning is accepted widely and proven to be accurate and efficient way of decision making. This paper discusses how Machine Learning techniques can be useful for student analysis purpose. Some of the most recent ML based solutions are discussed. We have also represented the student performance analysis related activities and their possible solution designs.

KEYWORDS:

Machine Learning, Students' Performance, Prediction, Classification

1. INTRODUCTION:

ML-Machine Learning makes computers learn the way human do. It is a branch of computer science with techniques to analyze the existing data, find hidden structure and relationships to derive the decision making model and use the decision making model to predict the future. ML explores the existing dataset called the training data to derive a model called the learner. The learner is tested with the existing dataset for its accuracy and efficiency analysis. Once a learner is found satisfactorily accurate and efficient, it is used for the future prediction.

Over the years, the performance analysis of students has become an essential activity to be performed by the academicians, government and recruiters. All these people need to find the necessary information of students' performance for the overall improvements in today's society. Different analysis could be prediction of results, selection of branch, curriculum gap, interests in subjects, technical and communication skills, suitability for recruitment etc. At the same time,

such analysis help to find the effectiveness of teaching in terms of success of teachers, quality of education of various institutes etc.

This paper discusses how ML-Machine Learning techniques could be useful for the students' performance analysis. This paper also discusses the fundamental types of Machine Learning techniques and their suitability for various types of students' performance analysis. Section 2 discussed fundamentals of ML. Section 3 discusses some existing ML based solutions of ML based students' performance analysis. Section 4 discusses some proposed solutions design.

2. MACHINE LEARNING

ML-Machine Learning [1,2,3] is a set of techniques which made computers learn the way human do. ML techniques are classified into various types according to the way they learn. Figure 1 shows a general model of ML process. ML based design starts with gathering data from various sources to build the training dataset. The next step would be the cleaning of data to avoid errors, duplications, partial or ambiguous data which may affect quality of learning. The next step would be the selection of the right – most appropriate ML algorithm for the model building. The next step is about analyzing the performance of built model for its accuracy and efficiency. The last step would be to use the built model for the new data and to transform results into visual representations such as graphs etc.

Figure – 1 – ML Process

ML algorithms depend on the way learning happens. They are primarily classified according to the fundamental way of learning: Supervised Learning, Unsupervised Learning and Reinforcement Learning. Figure – 2 shows these three types of ML and their fundamental differences.
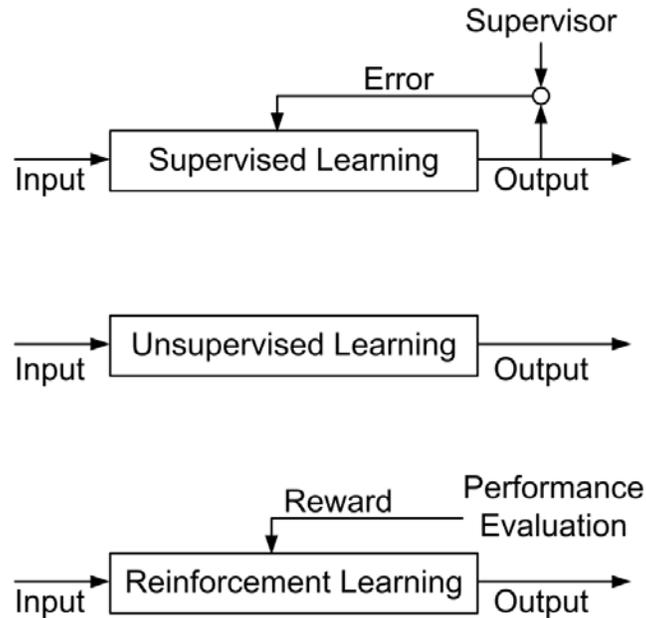
Figure – 2 – ML Types

Supervised Learning needs a training dataset to learn. The training dataset is prepared based on history, experiments, observations, records etc. This training dataset needs to be accurate and complete so all possibilities of the application could be covered. The Supervised Learning algorithm analyzes the training dataset and builds a model. Later on this model is evaluated with a subset of training dataset to measure its accuracy and efficiency. This type of learning is used mainly when the training dataset is possible to arrange. This type of learning is used for the prediction purpose in terms of classification or regression. Here supervisor is present in terms of the labeled data. The training dataset has all information of labeling of data and the learner needs to learn the relationships only [1,2,3].

Unsupervised Learning explores the data and derives the hidden relationships itself. The training data has no labels and the algorithm needs to find out the structure according to the values of various parameters. Such algorithms help in understanding complex systems with no historical details of the relationships of the elements. This learning suits well for the clustering purpose [1,2,3].

Reinforcement Learning suits well to adapt the real world learning performed by the human and other animals. The learning happens run-time without any prior processing. It is the type of learning which is based on trial and learn. The learner keeps interacting with the environment to learn. The environment responds with a reward. The reward is used by the learner to identify how accurate or effective the previous action was. The learner keeps learning and the decision making policy keeps getting improved [1,2,3].

3. LITERATURE REVIEW

Over the years researchers have proposed various ML-Machine Learning based solutions for various activities related with the students. This section discusses most recent of these solutions along with the type of student activities they address. A next term course recommendation system is proposed to help students while selecting next term course. The objective of the model is to predict what is the probability some particular student has of passing all of the courses he wants to enroll in the next term. long-short term memory LSTM is an effective and scalable type of RNN which is used for the prediction [4]. A framework for predicting student performance based on Machine Learning algorithm at H.E.K high school in Morocco is proposed. The modeling phase was developed in two stages: (1) model for forecasting the second semester results and (2) model for forecasting the results of the national exam. The results of these two models are used to predict the grade of the Bac for each student. This approach used multiple regression algorithm for the prediction [5]. An overview of machine learning in education with the focus on techniques for student dropout prediction is discussed in [6]. They have discussed the data imbalanced issue present in many of the existing solutions. Feature extraction for classifying students based on their academic performance is proposed in [7]. This approach accurately identifies students that are at risk. These students might fail the class, drop it, or perform worst than they usually do. Gradient Boosting and Random Forest classifiers. A Support Vector Regression for GPA Prediction is proposed in [8]. The features used in predicting GPA are semester 1 and semester 1 IP grades. The process of GPA prediction uses SVM regression, Linear Regression, and Simple Linear Regression. The performance of students' is predicted with deep learning based data transformation and factor analysis in [9]. In this research work, researchers have used different data preprocessing techniques and deep learning methods. A practical model for k-12 education is proposed in [10]. Three classifiers are used to develop these predictive models, including linear regression, decision tree, and Naive Bayes techniques. The objective of this model is to provide a method for identifying the most indicative features of a successful performance of students in grades K-12. A novel approach to measure students' cognitive skills is proposed with a multi-layer prediction solution in [11]. The student characteristic model is designed with various schedules of travel time, study time, family relationship, outing time, first group free time, second group free time. Intelligent Tutoring System enables students to construct individually realistic goals based on their current position. Recent developments in ITS have enabled customized education by suggesting optimal strategies for individual students to approach studying. A Deep Attentive Score Prediction model is prescribed to improve student engagement in [12]. Assessment Modeling, a class of fundamental pre-training tasks for general Interactive Educational Systems is proposed in [13]. A Comparison between content-based and interaction-based approaches is also given. A ML based support solution for students for academic decisions is proposed in [14]. This work is a holistic framework for assisting students in their decision throughout their entire study journey, and not only at one point of their study. This research tests and compares the performance of Decision

Trees, Random Forests, Gradient-Boosted trees, and Deep Learning machine learning regression algorithms to predict student GPA. An Apache Spark-based Platform for Predicting student performance is proposed in [15]. This research work addresses the limitations of conventional ML based solutions to predict student performance due to dependency on datasets. This solutions builds a distributed platform on Spark to predict missing grades of elective courses for undergraduate students.

4. PROPOSED SOLUTION DESIGN

This section discusses some of the proposed solution designs in which researchers may work by applying appropriate ML techniques. Most of these requirements are either traditionally being solved or using conventional data processing systems. These are classified into three categories. We have classified the performance analysis of students and their related activities in four main categories as shown below.
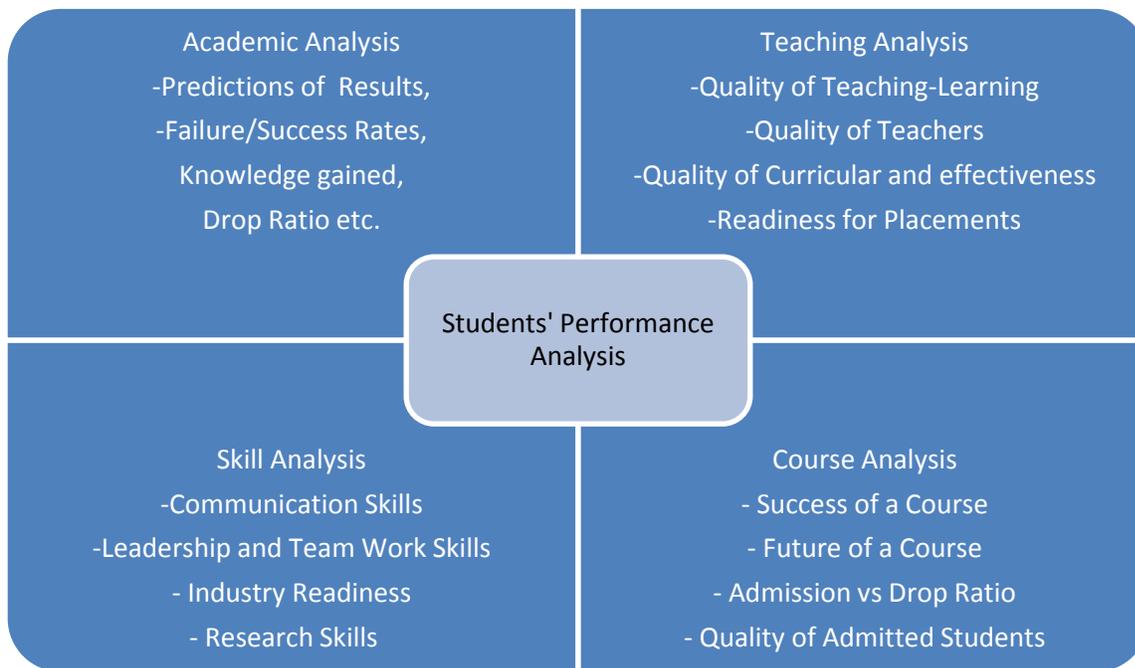
a. Academic Analysis: This group of activities is mainly related with the analysis of students' academic performance. One very obvious analysis which is widely done by the researchers in about prediction of performance of a student in next examination. Along with it, the failure or success rate of a class or a batch is also required to be done. In outcome based education, analysis of academic performance vs knowledge gained is essential. The drop ratio is also helpful in identifying the loopholes of curriculum.

b. Teaching Analysis: This group of activities is mainly related with the analysis of quality of teaching-learning of an institute / university. The quality and efforts of teachers need to be analyzed for their improvement as well as for their promotions. Quality of curriculum and readiness for placements are other parameters which could be determined through teaching analysis.

c. Skill Analysis: This group of activities is mainly related with the analysis of students' overall improvements in various skills. This group of activities are essential specially in the professional courses such as engineering. Communication, leadership, team work, research skills need to be analyzed so students' perform well after completion of respective course.

d. Course Analysis: This group of activities is mainly related with the analysis of students performance and its impact on a course. This activity help in identifying future admission rate of a course, requirements of up gradation of syllabus or closing a course.

4. CONCLUSION

In the era of computerization, machine learning is being applied almost everywhere. Education data mining is also enhanced by applying machine learning techniques for accurate and efficient decision making. This paper discusses the fundamentals of machine learning and how it could be used for the students performance prediction. We have discussed some of the most recent ML based approaches. Further to the discussion, we have presented students' performance analysis requirements into four groups. Each of the activities in each of this group could be tried to be solved by applying appropriate ML techniques. This paper could be helpful to the beginners who want to research in this direction.

REFERENCES

1. Jiawei Han, Micheline Kamber, and Jian Pei. Data mining: Concepts and techniques. Morgan Kaufmann Publishers Inc., 2011.

2. S. Wang, W. Chaovalitwongse, and R. Babuska. Machine learning algorithms in bipedal robot control. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(5):728--743, 2012.

3.   Ethem Alpaydin. 2010. Introduction to Machine Learning (2nd. ed.). The MIT Press.

4.   Volk, N. A., Rojas, G., & Vitali, M. V. (2020). UniNet: Next Term Course recommendation using Deep Learning, (September), 377–380. https://doi.org/10.1109/icacsis51025.2020.9263144

5.   Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammass, D. (2019). A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, *24*(6), 3577–3589. https://doi.org/10.1007/s10639-019-09946-8

6   Prenkaj, B., Velardi, P., Stilo, G., Distante, D., & Faralli, S. (2020). A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. *ACM Computing Surveys*, *53*(3). https://doi.org/10.1145/3388792

7   Polyzou, A., & Karypis, G. (2018). Feature extraction for classifying students based on their academic performance. *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*, 356–362.

8   Dewi, K. E., & Widiastuti, N. I. (2020). Support Vector Regression for GPA Prediction. *IOP Conference Series: Materials Science and Engineering*, *879*(1). https://doi.org/10.1088/1757-899X/879/1/012112

9   Dien, T. T., Luu, S. H., Thanh-Hai, N., & Thai-Nghe, N. (2020). Deep learning with data transformation and factor analysis for student performance prediction. *International Journal of Advanced Computer Science and Applications*, *11*(8), 711–721. https://doi.org/10.14569/IJACSA.2020.0110886

10   Harvey, J. L., & Kumar, S. A. P. (2019). A Practical Model for Educators to Predict Student Performance in K-12 Education using Machine Learning. *2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019*, 3004–3011. https://doi.org/10.1109/SSCI44817.2019.9003147

11   Ahmad, S., Li, K., Amin, A., Anwar, M. S., & Khan, W. (2018). A multilayer prediction approach for the student cognitive skills measurement. *IEEE Access*, *6*, 57470–57484. https://doi.org/10.1109/ACCESS.2018.2873608

12   Lee, Y., Kim, B., Shin, D., Kim, J. H., Baek, J., Lee, J., & Choi, Y. (2020). Prescribing deep attentive score prediction attracts improved student engagement. *ArXiv*, 3–8.

13   Choi, Y., Lee, Y., Cho, J., Baek, J., Shin, D., Lee, S., … Heo, J. (2019). Assessment modeling: Fundamental pre-training tasks for interactive educational systems. *ArXiv*.

14   Fateh Allah, A. G. (2020). Using machine learning to support students′ academic decisions. *Journal of Theoretical and Applied Information Technology*, *98*(18), 3778–3796.

15   Mai, T. Le, Do, P. T., Chung, M. T., & Thoai, N. (2019). An apache spark-based platform for predicting the performance of undergraduate students. *Proceedings - 21st IEEE International Conference on High Performance Computing and Communications, 17th IEEE International Conference on Smart City and 5th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2019*, (August), 191–199. https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00041