TRACKING AND PREDICTING STUDENT PERFORMANCE USING MACHINE LEARNING

Bagal Rupali[1], Matele Bhavika[2], Wable Pragati[3]

[1,2,3]*Student, Dept. of Computer Engineering, JCOE Kuran , Maharashtra, India*

Abstract - *Educational Data Mining (EDM) and Learning Analytics (LA) research have emerged as interesting areas of research, which are unfolding useful knowledge from educational databases for many purposes such as predicting student's success. The ability to predict a student's performance can be beneficial for actions in modern educational systems. Existing methods have used features which are mostly related to academic performance, family income and family assets; while features belonging to family expenditures and students personal information are usually ignored. In this paper, an effort is made to investigate aforementioned feature sets by collecting the scholarship holding students data from different universities. Learning analytics, discriminative and generative classification models are applied to predict whether a student will be able to complete his degree or not. Experimental results show that proposed method significantly outperforms existing methods due to exploitation of family expenditures and students personal information feature sets.*

Accurately predicting students future performance based on their ongoing academic records is crucial for effectively carrying out necessary pedagogical interventions to ensure students on-time and satisfactory graduation, predicting student performance in completing degrees (e.g. college programs) is much less studied and faces new challenges

Key Words:
Data Mining, Machine Learning, Personalized Education, Tracking Students Performance, Course Prediction, and Recommendation System

1. INTRODUCTION

To address the aforementioned challenges, we proposed a novel algorithm for predicting student's performance in college programs given his/her current academic records. In Proposed studies shows that academic performances of the students are primarily dependent on their past performances. Our investigation confirms that past performances have indeed got a significant influence over students' performance. Further, we confirmed that the performance of SVM increases with increase in dataset size. System will comprise the tracking of detailed information of a student regarding his academics and curricular activity and would predict the right learning Courses using an algorithm over the information tracked meeting the ambition or the goal for a student. In the last decade, school conducts examination manually. It has so many problems. The existing systems are very time consuming. It is difficult to analyze the exam manually. Results are not precise as calculation and evaluations are done manually. Result processing after summation of exam takes more time as it is done manually. So we introduce a Preschool examination Portal system, which is fully computerized. Existing system is a large man power process and is difficult to implement. It provides an easy to use environment for both Test Conductors and Students appearing for Examination. The main objective of Preschool examination Portal system is to provide all the features that an Examination System must have, with the" interfaces that don't Scare it's Users!"

1.1 Related Work

Nguyen Thai-Nghe, Lucas Drumond, Tom a s Horv ath, and Lars, Schmidt-Thieme, University of Hildesheim, "MultiRelational Factorization Models for Predicting Student Performance". [1] In this paper we propose to exploit such multiple relationships by using multi-relational MF methods. Experiments on three large datasets show that the proposed approach can improve the prediction results. Predicting student performance (PSP) is the problem of predicting how well a student will perform on a given task. It has gained more attention from the educational data mining community recently. Previous works show that good results can be achieved by casting the PSP to rating prediction problem in recommender systems, where students, tasks and performance scores are mapped to users, items and ratings respectively. Yiding Liu1 TuanAnh Nguyen Pham2 Gao Cong3 Quan Yuan describe the An Experimental Evaluation of Point of interest Recommendation in Location based Social Networks-2017 In this paper, we provide an all-around Evaluation of 12 state-of-the-art POI recommendation models. From the evaluation, we obtain several important findings, based on which we can better understand and utilize POI recommendation Models in various scenarios [2].

Beverly Park Woolf, Ryan Baker, Worcester Polytechnic, Institute Erwin P. Gianchandani, "From Data to Knowledge to Action: Enabling Personalized Education". [2] We describe how data analytics approaches have the potential to dramatically advance instruction for every student and to enhance the way we educate our children. The Internet, intelligent environments, and rich interfaces (including sensors) allow us to capture much more data about learners than ever before and the quantities of data are growing at a rapidly accelerating rate.

Nazeema Alli, Rahim Rajan, and Greg Ratliff, "How personalized learning unlocks student Success" [3] EDUCAUSE Review is the general-interest, bimonthly magazine published by EDUCAUSE. With a print publication base of 22,000, EDUCAUSE Review is sent to EDUCAUSE member representatives as well as to presidents/chancellors, senior academic and administrative leaders, non-IT staff, faculty in all disciplines, librarians, and corporations. It takes a broad look at current developments and trends in information technology, what these mean for higher education, and how they may affect the college/university as a whole.

Personalized Grade Prediction: A Data Mining Approach [4] Proposes an algorithm that predicts the final grade of each student in a class. It issues a prediction for each student individually, when the expected accuracy of the prediction is sufficient. The algorithm learns online what is the optimal prediction and time to issue a prediction based on past history of students performance in a course.

1.2Pr0posed Approches

In proposed to predict the performance of students in an academic organization, Present studies shows that academic performances of the students are primarily dependent on their past performances. Our investigation confirms that past performances have indeed got a significant influence over students' performance. Further, we confirmed that the performance of neural networks increases with increase in dataset size. Additionally, this work will also impact curriculum design in degree programs and education policy design in general. Future work includes Extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

Proposed Architecture

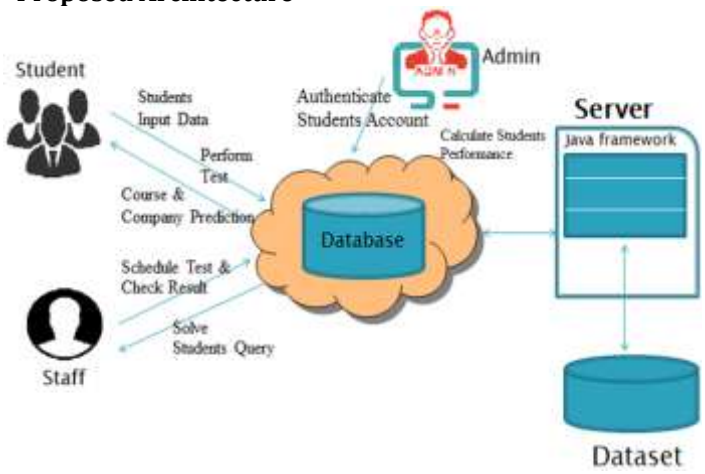

Fig-1 *Proposed Architecture*

1.3Algorithm

1. **k-Means Clustering:**

k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given dataset through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the scheme, and that the sparse configuration and rank one significantly improves the performance of the recommendation. Better choice is to place the as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

2. **SVM:**

In data analytics or decision sciences most of the time we come across the situations where we need to classify our data based on a certain dependent variable. To support the solution for this need there are multiple techniques which can be applied; Logistic Regression, Random Forest Algorithm, Bayesian Algorithm are a few to name. SVM is a machine learning technique to separate data which tries to maximize the gap between the categories.

2. CONCLUSION

Present studies shows that academic performances of the students are primarily dependent on their past performances. Our investigation confirms that past performances have indeed got a significant influence over students' performance. Further, we confirmed that the performance of SVM, K-Means increases with increase in dataset size. Machine learning has come far from its nascent stages, and can prove to be a powerful tool in academia. In the future, applications similar to the one developed, as well as any improvements thereof may become an integrated part of every academic institution.

Additionally, this work will also impact curriculum design in degree programs and education policy design in general. Future work includes extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

3. RESULT ANALYSIS:

In this work will also impact curriculum design in degree programs and educational policy design in general. In this project, we proposed a novel algorithm for predicting student's performance in college programs given his/her current academic records. Our data-driven approach can be used together with other pedagogical methods for evaluating student's performance and provide valuable information for academic advisors to recommend subsequent courses to students and carry out educational interventions measures if necessary.

REFERENCES

[1] The White House, Making college affordable, https://www:whitehouse:gov/issues/education/higher-education/making-collegeaffordable, 2016.
[2] The White House, Making college affordable, https://www:whitehouse:gov/issues/education/higher-education/making-collegeaffordable, 2016.
[3] H.Cen,K.Koedinger,andB.Junker,Learning factors analysis a general method for cognitive model evaluation and improvement, in International Conference on Intelligent Tutoring Systems. Springer,2006, pp. 164175.
[4] M. Feng, N. Heffernan, and K. Koedinger, Addressing the assessment challenge with an online system that tutors as it assesses, User Modelingand User Adapted Interaction, vol. 19, no. 3, pp. 243266, 2009.
[5] H.-F. Yu, H.-Y. Lo, H.-P. Hsieh, J.-K.Lou, T. G. McKenzie, J.-W.Chou, P.H.Chung,C.-H.Ho,C.-F.Chang,Y.-H.Weietal.,Feature engineering and classifier ensemble for kdd cup 2010, in Proceedingsof the KDD Cup 2010 Workshop, 2010, pp. 116.
[6] Z. A. Pardos and N. T. Heffernan, Using hmms and bagged decision trees to leveragerich features of user and skill from an intelligent tutoring system dataset, Journal of Machine Learning Research W CP, 2010.
[7] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, Personalized grade prediction: A data mining approach,in Data Mining(ICDM),2015IEEEInternational Conference on. IEEE, 2015, pp. 907912.
[8] C. G. Brinton and M. Chiang, Mooc performance prediction via click stream data and social learning networks,in2015IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 22992307.
[9] KDD Cup, Educational data minding challenge, https://pslcdatashop:web: cmu:edu/KDDCup/, 2010.13
[10] C. Marquez-Vera, C. Romero, and S. Ventura, Predicting school failure using data mining, in Educational Data Mining 2011, 2010.
[11] Y.-h. WangandH.-C.Liao,Data mining for adaptive learning in a tesl based elearning system,Expert Systems with Applications,vol. 38,no. 6,pp. 64806485,
[12] N.Thai-Nghe,L.Drumond,T.Horvath,L.Schmidt-Thiemeetal.,Multi-relational factorization models for predicting student performance ,in Proc. of the KDD Workshop on Knowledge Discovery in Educational Data.Citeseer, 2011.