Tracking and Predicting Student Performance Using Machine Learning

Bagal Rupali Ramdas
Department of Computer Engineering,jcoe
University of Pune
bagalrupali26@gmail.com

Matele Bhavika Machhindra
Department of Computer Engineering,jcoe
University of Pune
matelebhavika9@gmail.com

Wable Pragati Kailas
Department of Computer Engineering,jcoe
University of Pune
wablepragati10@gmail.com

Prof. Raut M. V.
Department of Computer Engineering,jcoe
University of Pune
raut.manju23@gmail.com

***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 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.**

***Index Terms*-Data Mining, Machine Learning, Tracking Students Performance, Course Prediction, and Recommendation System.**

I. INTRODUCTION

In proposed studies shows that academic performances of the primarily dependent on their past performances. Our investigation confirms that past performance have indeed got a significant influence over student's performance. Further, we confirmed that the performance of SVM increases with increases in dataset size.

System will comprise the tracking of detailed information of a student regarding his academic 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 analyse 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 Pre-school 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. In this paper, a model is proposed to predict the performance of students in an academic organization. The algorithm employed is a machine learning techniques called neural network. Further, the importance of several different attribute, or features is considered, in order to determine which of these are correlated with student performance.

II. RELATED RESEARCH

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.

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.

III.IMPLEMENTATION

A.Algorithm used for Prediction System in Machine Learning:

1.Naïve Bayes Classifier Algorithm:

Naive Bayes is a kind of classifier which uses the Bayes Theorem.[5] It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class.Naive Bayes classifier gives great results when we use it for textual data analysis. Since it is a probabilistic model, the algorithm can be coded up easily and the predictions made real quick. Real-time quick. Because of this, it is easily scalable and is traditionally the algorithm of choice for real-world application that are required to respond to user's requests instantaneously.Using this algorithm, we will be dealing with the probability distributions of the variables in the dataset and predicting the probability of the response variable belonging to a particular value, given the the attributes of a new instance.

B.Algorithm Used for Classification:

1.SVM (Support Vector Machine) Algorithm:

The support-vector machine algorithm, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications. Support vectors are simply the co-ordinates of individual observation. Support vector machines is a supervised machine learning algorithm which can be used for both classification or regression challenges. It is mostly used for classification problems.[6] SVMs are based on the idea of finding a hyperplane that divides a dataset into two classes. Support vectors are the data points nearest to the hyperplane, the points of a dataset that, if removed, would alter the position of the dividing hyperplane.

C. Algorithm used for Chatbot Model:

1. Sentiment Analysis:

Sentiment Analysis is an automated process that programmatically extracts topics from texts and the feeling of the writers towards such topics.[7] Chatbots are meant to automate and streamline communication with users, providing immediate responses and improving the customer experience through a friendly, conversational interface. One very common use of chatbots is first level customer support, through, goes far beyond the actual help provided to users. It is a valuable source of information about products and services.

IV.OVERVIEW OF PROPOSED MODEL

A. Problem Statement:

System will comprise the tracking of detailed information of a student regarding his academic and curricular activity and would predict the right learning courses using algorithm over the information tracked meeting the ambition or the goal for a student.

B. Objectives of model:

To develop a system for automatic identification the basic goals and ambitions of the student. Applying various kinds of techniques to effectively calculate students performance and recommend course according to the result. To track students performance and motivate him/her towards their goals and providing them with a right path.

C. Proposed Model:

The main purpose behind the proposed system to provide a comprehensive computerized system, which can capture, collate and analyse the data from these wards and evaluate the impact of the program. The proposed method has two major features, First, a bi-layered structure comprising of multiple base, predictors and a cascade of ensemble predictors is developed for making predictions based on student's probabilistic matrix factorization is proposed to discover course relevance, which is important for constructing efficient base predictors.

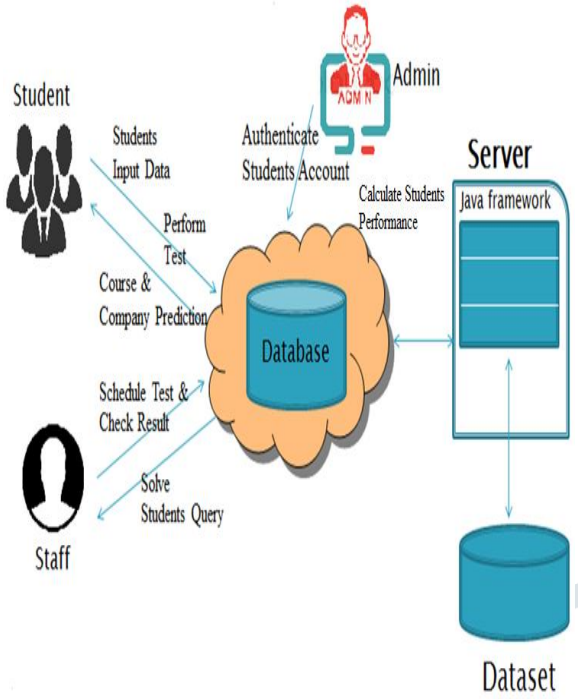

Figure 1. System block diagram

First students can differ tremendously in terms of backgrounds as well as their chosen areas, resulting in different selected courses as well as sequences in which they take the courses. Second, students may take many courses but not all courses are equally informative for predicting student's future performance. Utilizing the students past performance in all courses that he/she has completed not only increases complexity but also introduces noise in the prediction, there by degrading the prediction performance. Third, predicting student performance in a degree program, it does tracking and updating as the student finishes new courses over time.

V. Conclusion

Present studies shows that academic performances of the students are primarilydependent on their past performances. Our investigation confirms that past performanceshave indeed got a significant influence over students' performance. Further,we confirmed that the performance of neural networks increases with increase in
dataset size. In the future, applications similar to the one developed,as well as any improvements thereof may become an integrated part ofevery academic institution. This project can be used in any organization, college as
analysis purpose.

VI. Result

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 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 pedagogical interventions measures if necessary.

ACKNOWLEDEGMENT

Authors want to acknowledge Principal, Head of department and guide of their project for all the support and help rendered. To express profound feeling of appreciation to their regarded guardians for giving the motivation required to he finishing of paper.
.

REFERENCES

[1] H. Cen, K. Koedinger, and B. Junker, Learning factors analysis a general method for cognitive model evaluation and improvement, in International Conference on Intelligent Tutoring Systems, Springer, 2006, pp. 164175

[2]M. Feng, N. Heffernan, and K. Koedinger, Addressing the assessment challenge with an online system that tutors as it assesses, User Modelling and User-Adapted Interaction, vol. 19, no. 3, pp. 243266, 2009.

[3] 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.Wei et al., Feature engineering and classifier ensemble for kdd cup 2010, in Proceedings of the KDD Cup 2010 Workshop, 2010, pp. 116

[4] Z. A. Pardos and N. T. Heffernan, Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset, Journal of Machine Learning Research W CP, 2010.

[5] http://dataaspirant.com/2017/02/06/naive-bayes-classifier-machine-learning

[6]www.analyitcsvidhya.com/blog/2017/09/understanding-support-vector-machine-example-code/

[7]https://chatbotsmagazine.com/advanced=chatbot-analytics-and-sentiment-analysispart-1-17a8e764d7e1

[8] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, Personalized grade prediction: A data mining approach, in Data
Mining (ICDM), 2015IEEE International Conference on. IEEE, 2015, pp. 907912.

[9] C. G. Brinton and M. Chiang, Mooc performance prediction via click stream data and social learning networks, in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 22992307.

[10] C. Marquez-Vera, C. Romero, and S. Ventura, Predicting school failure using data mining, in Educational DataMining 2011, 2010