



IMPROVED LATENT PREDICTOR MODEL ON LOGATHEMIC

FEATURE FOR STUDENTS PERFORMANCE

Mr.G.Bhanuprasad¹, B.Sravani², D.Deepika³, B.Sai Sruthi⁴

¹Associate Professor, Department of CSE, Malla Reddy Engineering College for Women,
Hyderabad, TS, India

^{2,3,4} UG Scholar, Department of CSE, CSE, Malla Reddy Engineering College for Women, Hyderabad,
Telangana, India

ABSTRACT:

Towards automation to do mundane tasks and the expectations for students already equipped with good programming skills is on the rise. In parallel, there have been a rising number of students who find it difficult to attain the skills necessary in order to get the dream IT job they desire. The aim of this project is to bridge the gap between the employer and the future employee of the company by the use of SPAS at college level. Student performance analysis system (SPAS) is an online web application system which enables students to know prior hand if their level of skills for the placement is enough to get placed or not, given the necessary inputs. SPAS have an intelligent learning algorithm which utilizes a rich database, analyses the records of previous students' traits and develops a model for further prediction. The performance evaluation of students by SPAS is by the cumulative predictor algorithm involving generation of several random forest trees on the available data. SPAS learn and create its model reaching higher accuracy with increasing data availability

Keywords: SPAS, EDM, IT, DM, training.

1. INTRODUCTION:

Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. There are several

data regarding the students which stay unused with untapped potential of data mining which could revolution is the field of education. Since the ultimate aim of an educational institution is to create a pool of skilled professionals to take on the society to a next upgraded level, they need to create an environment for their students to grow



in every vertical by giving them right exposure and training. Most of the educational institutions, maintain huge databases of students and the information keeps on increasing with time, but there is no action taken to gain knowledge from it. DM has the suitable techniques in mining the data to discover new information and knowledge about students. DM provides various methods for analysis which include classification, clustering, and association rules. Classification, one of the prediction algorithms, classifies the data (constructs a pattern) based on the training set and uses the pattern to classify a new data (testing set). IHL faces a major challenge in order to improve and manage the organization to be more efficient in managing students' activities. To achieve this target, DM is considered as the one of most suitable technique in giving additional insights to the IHL community to help them make better decisions in educational activities. The IHL make use of WEKA tool in order to build a model and predict the SAP in order for the professors to provide the students with individual attention. InSAP system, the classification method is selected to be applied on the students' data.

This pattern will be used to improve the SAP and to overcome the issues of low grades obtained by students. In our proposed system, student performance analysis system (SPAS) provides

means for students to get an idea regarding their profile is suitable for getting placements, for which we make a clear cut analysis of all the data parameters required for the classification and the students' prediction is made based on the collected data. For this, the placement details of the CSE passed out students of the batch 2016 and 2017 of our institution is considered. The data is collected from the concerned and it is normalized. The unnecessary attributes for prediction are removed and the required parameter data is cleaned and converted into Raff format for the analysis in WEKA tool. Various existing algorithms are applied in the WEKA tool and the most efficient algorithm is considered for improvement to be used in SPAS. The cumulative predictor algorithm builds a prediction model on the previous years' student data which can be applied on future data sets. This algorithm is found to be more efficient than the existing systems and this model built is used in SPAS. SPAS provide an interface in a web platform which enables both teachers as well as students to predict the outcome of placements based on the model's results. Data provided to SPAS can also be designed to predict even more information like possible arrear students, product based potential students and much more. Since there is support from the institution, a data analysis to determine the necessary parameters and the collection of



information for the parameters can be done from the students and the institution. The existing model focus mainly on probable low graders, so that the professors can provide individual attention, whereas the proposed system allows both students and staffs who can work mutually in order to increase the placements in the college.

2. LITERATURE SURVEY

The prediction of students' academic performance using classification data mining techniques

AUTHORS: Ahmad, F., Ismail, N.H. and Aziz, A.A Data Mining provides powerful techniques for various fields including education. The research in the educational field is rapidly increasing due to the massive amount of students 'data which can be used to discover valuable pattern pertaining students' learning behavior. This paper proposes a framework for predicting students' academic performance of first year bachelor students in Computer Science course. The data were collected from 8 year period intakes from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records, and family background information. Decision Tree, Naive Bayes, and Rule Based classification techniques are applied to the students' data in order to produce the best students' academic performance prediction model. The experiment result shows

the Rule Based is a best model among the other techniques by receiving the highest accuracy value of 71.3%. The extracted knowledge from prediction model will be used to identify and profile the student to determine the students' level of success in the first semester.

Predicting student performance by using data mining methods for classification

AUTHORS: Dorina Kabakchieva Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making. This paper presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management.

A review on predicting student's performance using data mining techniques

AUTHORS: Shahiria, A.M., Husaina, W. and Rashida, N.A Predicting student's performance becomes more challenging due to the large volume of data in educational databases. Currently in Malaysia, the lack of existing system to analyze and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods



for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors affecting student's achievements in particular courses within Malaysian context. Therefore, a systematically literature review on predicting student performance by using data mining techniques is proposed to improve student's achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict student's performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data. We could actually improve student's achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

Classification and performance evaluation using data mining algorithms

AUTHORS: Vasani, V.P. and Gawali, R.D. In this paper, classification of the data collected from students of polytechnic institute has been discussed. This data is pre-processed to remove unwanted and less meaningful attributes. These students are then classified into different categories like brilliant, average, weak using decision tree and naïve Bayesian algorithms. The

processing is done using WEKA data mining tool. This paper also compares results of classification with respect to different performance parameters.

3. METHODOLOGY

EXISTING SYSTEM:

The data contains the students' demographics, previous academic records, and family background information. DT, NB, and Reclassification techniques are applied to the students' data in order to produce the best SAP prediction model. The experiment result shows the RB is a best model among the other techniques by receiving the highest accuracy value of 71.3%. The extracted knowledge from prediction model will be used to identify and profile the student to determine the students' level of success in the first semester. This project acts as the basis of SPAS and gives a clear idea regarding the parameters involved in predicting students' performance.

PROPOSED SYSTEM

In this paper, we have introduced new algorithm called cumulative predictor to predict student performance beforehand so they can know whether they are fit for IT dream job or not. We have built cumulative predictor algorithm on top of decision tree (J48) algorithm and in this

algorithm author has put cumulative predictor model generation in loop so the algorithm can choose best accuracy model with less error rate. In paper code we can see that the decision tree is in loop to get the best model. From the selected algorithm, we can see that the cumulative predictor is getting built with decision tree in for loop and selecting model with best accuracy.

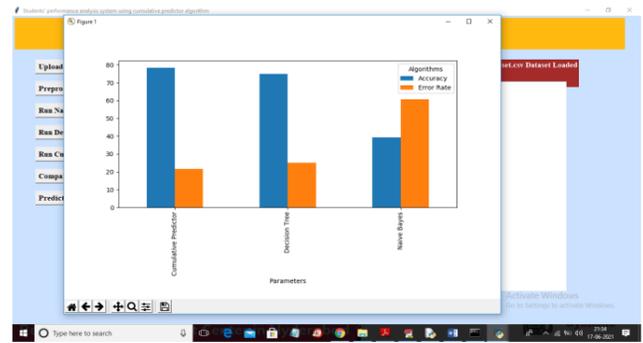
Module Description To implement this project we have designed following modules

- 1) Upload dataset: using this module we will upload dataset to application
- 2) Preprocess Dataset: dataset contains missing and non-numeric values so by applying pre-processing steps we are replacing missing values with 0 and converting non-numeric values into numeric code. For example NO will be replaced with 0 and YES will be replaced with 1.
- 3) Run Naive Bayes Algorithm: using this module we will train Naïve Bayes with above dataset and then calculate accuracy and error rate
- 4) Run Decision Tree Algorithm: using this module we will train Decision Tree with above dataset and then calculate accuracy and error rate
- 5) Run Cumulative Predictor Algorithm: using this module we will train Cumulative Predictor

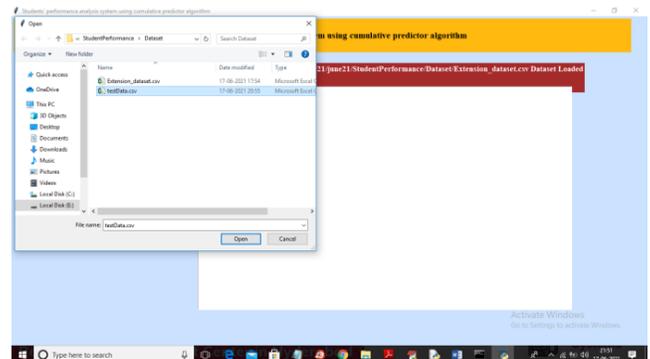
with above dataset and then calculate accuracy and error rate till we get best accuracy

6) Comparison Graph: using this module we will display comparison graph between all algorithm.

7) Predict Performance from Test Data: Using this module we will upload test data and then cumulative predictor will predict performance.



In above graph x-axis represents algorithm name and y-axis represents accuracy and error rate and in above algorithm we can see CP got high accuracy with less error rate. Now click on 'Predict Performance from Test Data' button to upload test dataset and to get prediction result.





- students', Proceedings of 10th Mexican International Conference on Artificial Intelligence, pp.169–174.
3. Jain, R. and Minz, S. (2008) 'Drawing conclusion from forest cover type data the hybridized rough set model', Journal of the Indian Society of Agricultural Statistics, Vol. 62, No. 1, pp.75–84.
4. Kabakchieva, D. (2013) 'Predicting student performance by using data mining methods for classification', Cybernetics and Information Technologies, Vol. 13, No. 1, pp.61–72 [online]
6. <http://dx.doi.org/10.2478/cait-2013-0006> (accessed 16 December 2016).
7. Mishra, T., Kumar, D. and Gupta, S. (2014) 'Mining students' data for performance prediction', Proceedings of Fourth International Conference on Advanced Computing and Communication Technologies, pp.255–262.
8. Oladokun, V.O., Adebajo, A.T. and Charles-Owaba, O.E. (2008) 'Predicting students' academic performance using artificial neural network: A case study of an engineering course', The Pacific Journal of Science and Technology, Vol. 9, No. 1, pp.72–79.
9. Patel, P.S. (2015) 'Various data mining techniques used to study student's academic performance', International Journal of Computer Science and Mobile Applications, June, Vol. 3, No. 6, pp.55–58.
10. Patidar, P., Dangra, J. and Rawar, M.K. (2015) 'Decision tree C4.5 algorithm and its enhanced approach for educational data mining', Engineering Universe for Scientific Research and Management.