

Students Performance Prediction in Online Courses Using Machine Learning Algorithms

G Mallikharjuna Rao¹, and Prof. K. Kiran Kumar²

¹M.Tech, Department of Computer Science

²Professor and HOD, Department of Computer Science

^{1,2}Chalapathi Institute of Engineering and Technology, Guntur, AP, India – 522034

Abstract— Automatic Student performance prediction is a crucial job due to the large volume of data in educational databases. This job is being addressed by educational data mining (EDM). EDM develop methods for discovering data that is derived from educational environment. These methods are used for understanding student and their learning environment. The educational institutions are often curious that how many students will be pass/fail for necessary arrangements. In previous studies, it has been observed that many researchers have intension on the selection of appropriate algorithm for just classification and ignores the solutions of the problems which comes during data mining phases such as data high dimensionality, class imbalance and classification error etc. Such types of problems reduced the accuracy of the model. Several well-known classification algorithms are applied in this domain existing models a student performance prediction model based on supervised learning decision tree classifier. In addition, an ensemble method is applied to improve the performance of the classifier. Ensemble methods approach is designed to solve classification, prediction problem. we propose a method for predicting final grades of students by a Recurrent Neural Network (RNN) from the log data stored in the educational systems. We applied this method to the log data from 108 students and examined the accuracy of prediction. From the experimental results, comparing with multiple regression analysis, it is confirmed that an RNN is effective to early prediction of final and suitable job for the student based on their academic performance and knowledge on skill set.

Keywords— machine learning, algorithms, RNN, data mining, recurrent neural network.

INTRODUCTION

Educational quality is compulsory in the development of each country. The data amount in education domain is getting increase day by day with the help of admission system, academic information system, learning management system, e-learning etc. The data collected from students are usually used for making simple quires for decision making. But most of the data remain unused due to complexity and large volume data sets. Therefore,

to analyze this huge amount of educational data is the great interest to predict student performance. Data mining is the practice of find out useful information Educational quality is compulsory in the development of each country. The data amount in education domain is getting increase day by day with the help of admission system, academic information system, learning management system, e-learning etc. The data collected from students are usually used for making simple quires for decision making. But most of the data remain unused due to complexity and large volume data sets. Therefore, to analyze this huge amount of educational data is the great interest to predict student performance. Data mining is the practice of find out useful information

In recent years, the use of ICT based educational systems has been widely spread. These systems enable us to collect many types of log data that corresponds to learning activities of students. By analyzing these logs using data mining techniques, we can determine learning patterns of students, which helps teachers in detecting “at-risk” students ([1]).

At Kyushu University, a learning support system called the M2B system was introduced in October 2014. The M2B system consists of three subsystems, the elearning system Moodle, the e-portfolio system Mahara, and the e-book system BookLooper provided by Kyocera Maruzen, Inc. Using the logs of these systems, a number of investigations have been conducted ([3], [4], [5]).

An early prediction of students’ final grades is an important task in the field of learning analytics, e.g., investigated in [6] using regression analysis. In this paper, we propose a method for predicting students’ final grades by a neural network approach, using the log data of the M2B system.

Particularly, in order to treat time series data of each week in a course, we use a variant of a neural network, called a Recurrent Neural Network (RNN) ([2]). By comparing our results with the result obtained using regression analysis, we show the performance of prediction of students’ final grades using RNN.

Existing Models:

Dorina et al. [1] proposed a predictive model for student's performance by classifying students into binary class (successful / unsuccessful). The proposed model was constructed under the CRISP-DM (Cross Industry Standard Process for Data Mining) research approach. The classification algorithms (OneR, J48, MLP and IBK) were applied on the given dataset. The results show that the highest accuracy was achieved by the MPL model (73.59%) for identification of successful while other three models perform better for the identification of unsuccessful students.

The model was unable to work out for data high dimensionality and class balancing problems. Edin Osmanbegovic et al. [2] builds a model to predict student academic success in a course by reducing data dimensionality problem.

Various machine learning classifiers such as NB, MLP and j48 were evaluated in this study. The result shows that the Naïve Bayes gained the highest accuracy 76.65%.

The proposed model not handles the class imbalance problem. Carlos et al. [3] addressed a student failure prediction model based on machine learning techniques to resolve the class imbalance and data dimensionality problems.

Ten classifiers were applied on dataset. The ICRM classifier achieved the highest accuracy 92.7% among others. Due to varying student's characteristics at each educational level, the performance of proposed model was not tested for other levels of education. Another EDM Challenge is to predict the drop-outs of the students from their courses [4].

Four data mining methods with six combinations of attributes were participated in this study. The result shows that the support vector machine model with the combination of the predictor variables was more accurate while classifying the data.

The inclusion of an attribute, earned grades of pre-requisite courses in the data set was the limitation of this study because it might be possible that during study of any course the student might have improved his knowledge of pre-requisite of this course. Ajay et al. [5] conducted study on the prediction of student performance.

The main contribution of the study was to introduce a new social factor called "CAT" which describes that in early times Indians were divided into four types of groups on the basis of their social status etc, which have

a direct effect on the student education. Four classifiers oneR, MLP, J48, and IB1 were applied on the data set. The results indicated that the IBI model was the highest accuracy (82%) achieved. Build an improved version of the ID3 model, which predicts the student academic performance [6]. The weakness of the ID3 model was its intension to select those attributes as a node which had more values. In a result generated tree was not efficient.

The proposed model overcomes such problem. Two output classes were produced by this model (Pass and Fail). The classifiers including J48, wID3 and Naïve Bayes were applied and results compared. The wID3 achieved high accuracy 93%. Alaa Khalaf et al. [7] proposed a model to predict student success performance in courses.

Three Decision Tree classifiers such as (J48, Hoeding tree, Reptree) were employed by this study. The highest accuracy 91.47 % was achieved by Reptree. The model was unable to work out for data high dimensionality and class balancing problems.

DechThammasiri et al. [8] proposed a model to provide early classification of poor academic performance of freshmen. Four classification methods with three balancing methods were applied to resolve class imbalance problem.

In results the combination of support vector machine and SMOTE achieved the 90.24% highest overall accuracy.

An early warning system was proposed to predict the student learning performances during an online course based on their learning portfolios data [9]. The results showed the approaches accompanied by time dependent variables had high accuracy than other approaches which were not included it.

The model was not tested on offline mode. The performance might be decreased in offline mode using time dependent attributes.

Mostly previous studies were assumed that the data mining algorithms performed well with only large data sets but this study supported that the data mining is also suitable for small datasets as well [10]. This research proposed a student success prediction model.

Proposed Implementation:

To address the common issues of above literature, review such as class imbalance, data hi-dimensionality and classification errors, this study has proposed a model which have following phases.

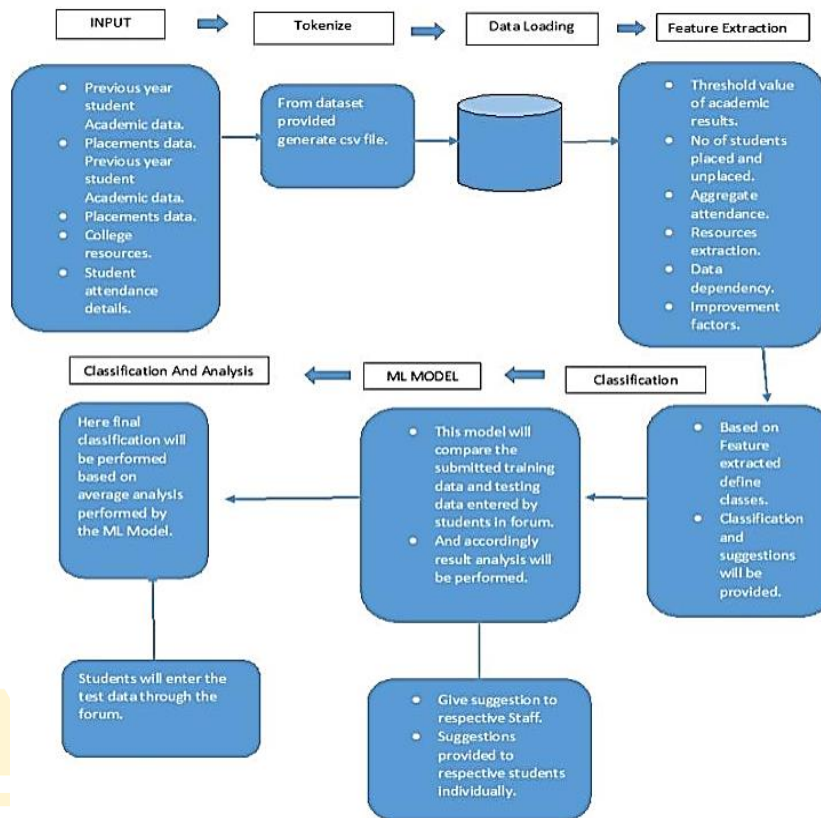


Fig. 1: System Architecture

Data Collection

We collected the learning logs of 108 students attending “Information Science” course, which started in April 2016. In this course, the teacher and students used the LMS, the e-portfolio system and the e-book system. The students were required each week to submit a report, to

answer a quiz, to write a logbook of a lecture, and to read slides for preview and review using the three systems. The logs of these learning activities were automatically graded by the system based on the criteria shown in Table 1:

Table 1: Criteria for grading learning activities

Activities	5	4	3	2	1	0
Attendance	Atten dance		Being late			absence
Quiz (rate of collect answer)	Above 80%	Above 60%	Above 40%	Above 20%	Above 10%	o.w.
Report	Submission		Late submission			No submission
Course views	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.
Slide views in Booklooper	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.
Markers in Booklooper	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.
Memos in Booklooper	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.
Actions in Booklooper	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.
Word count in Mahara	Upper 10%	Upper 20%	Upper 30%	Upper 40%	Upper 50%	o.w.

LITERATURE REVIEW

During my literature review study, to define the research problem as well as objectives, I would like to mention that I went through a lot of research papers and web application statistics. Few of paper's summarized details are as follows. In this first research study, author used feature selection technique to reduce the number of feature form the large attribute set. In this paper author use ASSISTments platform dataset which is a web based teaching system developed at Worcester Polytechnic institute and used with 4th to 10th grade math students. In this paper author used technique to remove irrelevant, redundant or noisy data. In this paper author used various classification algorithm and ranker algorithm to find top most contributed attribute and removed the less appropriate attribute. This helps to speeds up the process of data mining and improves its performance parameters such as predictive accuracy. [1] In this research paper author used three different approaches. Cross tabulation analysis, Feature selection and balancing imbalance data. Features selection method is used to select those attribute which are highly affected dependent variables. Classification tree is built considering all available attributes. This method finds out all possible splits that can occur for each indicator variable at each node. The search stops when the split with the largest imprudent in goodness of fit is found. A few element choice calculations are connected and includes positioning higher in numerous calculations are chosen. In this way 15 vital parameter are chosen from unique 77 attributes. Misbalancing issue is resolved by using data balancing and rebalancing algorithm specifically SMOTE (Synthetic Minority Over sampling technique). Ten fold cross validation is used for establishing training and testing data from original data. This data set is prepared in three categories. First category contains data with all 77 attributes. Next category contains data with 15 important attributes. Last category contains balanced data after applying rebalancing technique in weak. [2] This paper provides the hybrid approach for outlier detection. They used two algorithms: K-mean and Neural Network. The proposed method use Integrating Semantic Knowledge (SOF- Semantic outlier factor) for outlier detection. This method detects the semantic outlier. This technique identifies the semantic anomaly. Semantic exception is an information point that acts uniquely in contrast to other information focuses in a similar class or same bunch or cluster. The main motive of this research was to reduce the number of outliers in clusters as well as data by improving the cluster formulation methods so that outlier rate reduces. It also decreases the error and improves the accuracy. The result showed that the hybrid algorithm performs better than that of genetic k-means. This proposed strategy

manages content and date dataset that has not been executed before using genetic kmeans. [3] This research study describes the various approaches such as Neural network, K- Nearest Neighbour, Bayesian Classifier, Fuzzy Logic and decision tree classification Algorithms for implementation of intrusion Detection system. With the help of this paper, it is clear that the data mining methods are used to perform the intrusion detection system But this paper don't describe which technique is best for all of these. [4]

SYSTEM REQUIREMENTS

Hardware requirements:

- System : Pentium Dual Core.
- Hard Disk : 120 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 1GB.

Software requirements:

- Operating system : Windows 7.
- Coding Language : PYTHON

SYSTEM TESTING

The motivation behind testing is to find blunders. Testing is the way toward attempting to find each possible deficiency or shortcoming in a work item. It gives an approach to check the usefulness of parts, sub gatherings, congregations as well as a completed item It is the way toward practicing programming with the plan of guaranteeing that the Product framework lives up to its necessities and client desires and doesn't fizzle in an inadmissible way. There are different kinds of test. Each test type tends to a particular testing prerequisite.

Unit testing revolves affirmation effort around the humblest unit of Programming plan that is the module. Unit testing rehearses unequivocal routes in a module's control configuration to ensure absolute incorporation and most noteworthy error disclosure. This test bases on each module independently, ensuring that it limits fittingly as a unit. From now on, the naming is Unit Trying.

During this testing, each module is attempted only and the module interfaces are affirmed for the consistency with plan specific. Incredibly huge taking care of way are pursued for the ordinary results. All slip-up managing ways are furthermore attempted.

Fuse Testing: Incorporation testing watches out for the issues related with the twofold issues of check and

program advancement. After the item has been facilitated a lot of high solicitation tests are coordinated. The essential objective in this testing collaboration is to take unit attempted modules and develops a program structure that has been coordinated by plan.

Coming up next are the sorts of Reconciliation

Testing:

1. Top-Down Reconciliation

This procedure is a consistent method to manage the improvement of program structure. Modules are joined by moving dropping through the control reformist framework, beginning with the standard program module. The module subordinates to the standard program module are united into the plan in either a significance first or broadness first way.

In this strategy, the item is attempted from essential module and individual stubs are superseded when the test proceeds downwards.

2. Bottom-up Joining

This procedure begins the turn of events and testing with the modules at the most diminished level in the program structure. Since the modules are facilitated from the base up, dealing with required for modules subordinate to a given level is reliably available and the prerequisite for nails is murdered. The base up coordination technique may be executed with the going with propels:

- The low-level modules are joined into bunches into packs that perform a specific Programming sub-work.
- A driver (i.e.) the control program for testing is created to encourage explore data and yield.
- The pack is attempted.
- Drivers are taken out and bunches are joined moving upward in the program structure

The base up philosophies test each module freely and subsequently every module can't avoid being module is composed with a basic module and pursued for value.

Customer Acknowledgment Testing

Customer Acknowledgment of a structure is the basic factor for the accomplishment of any system. The system practical is gone after for customer affirmation by consistently remaining in contact with the approaching structure customers at the hour of making and making changes any spot required. The system made gives a friendly UI that can without a very remarkable stretch be seen even by a person who is new to the structure.

Yield Testing: In the wake of playing out the endorsement testing, the accompanying stage is yield attempting of the proposed structure, since no system could be useful if it doesn't make the vital yield in the foreordained game plan. Getting some data about the association required by them tests the yields made or appeared by the system practical. From now on the yield configuration is viewed as 2ly – one is on screen and another in printed plan.

Testing Procedure

A system for structure testing facilitates structure examinations and plan methodology into an overall orchestrated course of action of steps that results in the compelling improvement of programming. The testing framework must collaborate test orchestrating, analyze design, test execution, and the resultant data combination and evaluation. A procedure for programming testing ought to oblige low-level tests that are imperative to watch that a little source code area has been successfully executed similarly as evident level tests that favor huge system limits against customer necessities.

Programming testing is a fundamental segment of programming quality affirmation and addresses an authoritative review of detail plan and coding. Testing tends to a captivating peculiarity for the item. As such, a movement of testing are performed for the proposed system before the structure is ready for customer affirmation testing.

CONCLUSION

A deep neural network model is proposed in this paper for predicting the students' performance. It is the first time to use a deep neural network for the education data mining and predicting of students' performance. Through the experiment we found that a DNN can perform better even with less amount of data by having deep knowledge about dataset and quality tweak on the model. The proposed model achieved an accuracy of 84.3%. With larger dataset records and features, a DNN can achieve higher accuracy and will outperform other machine learning algorithm. This model is reliable and can help to predict a student's performance and identify students who has higher chance of failing before hand to provide remedy.

REFERENCES

[1] Livieris, et al. (2012): Predicting students' performance using artificial neural networks 8th PanHellenic Conference with International Participation Information and Communication Technologies, pp.321-328.

- [2] S. Kotsiantis, et al. (2003): Preventing student dropout in distance learning systems using machine learning techniques Applied Artificial Intelligence, 18(5), pp.411- 426.
- [3] Moucary, C.E., Khair, M. and Zakhem, W., 2011. Improving student's performance using data clustering and neural networks in foreign-language based higher education. The Research Bulletin of Jordan ACM, 2(3), pp 27-34
- [4] Yi, Hongsuk, Jung, H. and Bae, S., 2017, February. Deep Neural Networks for traffic flow prediction. In Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on (pp. 328-331). IEEE.
- [5] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553), pp. 436-444.
- [6] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T., 2014, November. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia pp. 675-678
- [7] Dahl, G.E., Yu, D., Deng, L. and Acero, A., 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on audio, speech, and language processing, 20(1), pp.30- 42.
- [8] Collobert, R. and Weston, J., 2008, July. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning pp.160-167
- [9] Koutina M, Kermanidis KL. Predicting postgraduate students' performance using machine learning techniques. In Artificial Intelligence Applications and Innovations 2011 (pp. 159-168). Springer, Berlin, Heidelberg.
- [10] Saini, P. and Jain, A.K., 2013. Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions. International Journal of Computer Applications, 84(14). Springer, Berlin, Heidelberg.
- [11] Agrawal, H. and Mavani, H., 2015. In Student Performance Prediction using Machine Learning.