Deep Regressor: Cross Subject Academic Performance Prediction System for University Level Students

B. Raveendran Pillai, GauthamJ

Abstract— ***Predicting the academic performance of students has been an important research topic in the Educational field. The main aim of a higher education institution is to provide quality education for students. One way to accomplish a higher level of quality of education is by predicting student's academic performance and there by taking earlyre- medial actions to improve the same. This paper presents a system which utilizes machine learning techniques to classify and predict the academic performance of the students at the right time before the drop out occurs. The system first accepts the performance parameters of the basic level courses which the student had already passed as these parameters also influence the further study. To pre- dict the performance of the current program, the system continuously accepts the academic performance parame- ters after each academic evaluation process. The system employs machine learning techniques to study the aca- demic performance of the students after each evaluation process. The system also learns the basic rules followed by the University for assessing the students. Based on the present performance of the students, the system classifies the students into different levels and identify the students at high risk. Earlier prediction can help the students to adopt suitable measures in advance to improve the per for- man ce. The systems can also identify the factor saffecting the performance of the same students which helps them to take remedial measures in advance.***

Keywords: machine learning, academic performance, deep learning, classification-regression methods, keras, tensorflow

I. INTRODUCTION

Many studies have been conducted in the e-learning field to analyze the academic performance of students and the early prediction of drop out in the academic courses. This would help the students take appropriate remedial steps before drop out occurs. In paper[1] the authors S. Kotsiants. C Pierrakeas and P Pintelas investigated the academic performance of the distance learning program for predicting dropout students. Machine earning techniques like Decision trees, Neural Networks, and Naıve Bayes algorithm are used. After the training and testing process, the performance is compared. Finally, it was concluded the Naıve Bayes algorithm found to be very efficient in the prediction of dropouts as it predicts with 83 percent accuracy. In the prediction process proposed by Lykourentzou, Giannoukos, Nikolopoulos and Loumosinpaper[2], the machine learning techniques; Support Vector Machines and feed forward neural networks had been used. For the training and testing purposes, both timeand time-invariant data were used. In the classification ofstu- dents based on performance using data mining technique is adopted.Continuous evaluation of student's performance of each subject can be done. This helpsthe Lecturerto take more attention to the student shavingpoor performance. A study using Machine learning[3] about the academic performance by Moseley and Mead forpre- dicting the dropouts of nursing students. In this case, a decision tree in which the Chi-square automated interac- tion detection was used. The dataset used in testing and validating phases are time-invariant (e.g. age, gender) and time-varying (e.g. grades, attendance). The system identifiedthestudentsatriskandthepredictionefficiency is 70 percent. The drawback of this method is that the prediction efficiency is only at a medium level and the method is not determining the academic performance but the system only finding out the dropoutstudents.

Revised Manuscript Received on September 22, 2019.

Prof.(Dr.)B. Raveendran Pillai Principal, PRS College of Engineering, Trivandrum, India. ravipoykali@gmail.com

GauthamJ, 2nd Year Mechanical Engineering, National Institute of Technology,Calicut, India. gauthamastro@gmail.com.

II. DEEP REGRESS OR MODEL

2.1 Problem Definition

In educational system some students even with a good academic background can also become weak due to many reasons; hardness in the subjects of study, poor teaching, and presentation method adopted in the educational in- situation, environmental conditions, lack of interest, poor health conditions, peer influence, etc. But if the educational institution has a system to consistently monitor the academic performance of the students and adopts proper remedial measures in time for the students with poor performances can be improved. But in a manual system, it is very difficult to consistently watch the academic activities of each student especially when the number of students is large. Also with the manual system, it is practically very difficult to predict accurately the dropout students based on the available data; the acquired performance valuation parameters up to the current stage. The machine learn- ing algorithms including deep learning can be employed for classifying the students based on the academic per for- mance and predict the dropouts more accurately.

Cross Subject Prediction Moreover it is difficult for a human to trace the relation between subjects infirst semester to subjects in second,third or any other semester,as the subjects are not same but different sub- jects of the same branch. There by making itdiffi cult to analyse a student for his performance in up- coming semester based on his performance in previous semesters.There are various models like GritNet[4]for which the input is time-variant data of the same subject that output is predicted for, Our model take in data of marks from some subjects in the previous semester and then predict for entirely different subjects.This is done by learning the student's knowledge level from different sources and this is one of the points that stand out for our model.

2.3 Architecture

Objective The main objective of this study is to de- velop a deep learning model to early predict more accu- rately the dropout of students in the 5th semester for an engineering degree course at the University. The input data are the performance parameters of the students of each subject in each semester up to the fourth semester.

Model The Deep Regressor model tries to predict the grade point for all the subjects in the output semester whichisacontinuously valuedoutputrangingfrom0to10. The corresponding grade points are converted to grades based on the grade criteria of the university using which we can predict whether a student will drop out ornot.

2.2 BaselineModel

In order to access how much value is brought by our model, we are comparing our model against the most ef- fective decision tree based regression algorithms avail- able.

XGBoostRegressor[6] This is an ensemble learn- ing method and it is a popular implementation of gradient boosting. One of the features which makes XGBoostst and out is the option to penalize complex model sthroughboth L1 and L2 regulariation. This preventso verfitting.

$$L_1 Loss = \sum_{i=1}^{n} (Y_i - \sum_{j=1}^{p} X_{ij} B)^2 + \lambda \sum_{j=1}^{p} B^2$$

$$L_2 Loss = \sum_{i=1}^{n} (Y_i - \sum_{j=1}^{p} X_{ij} B)^2 + \lambda \sum_{j=1}^{p} |B|$$

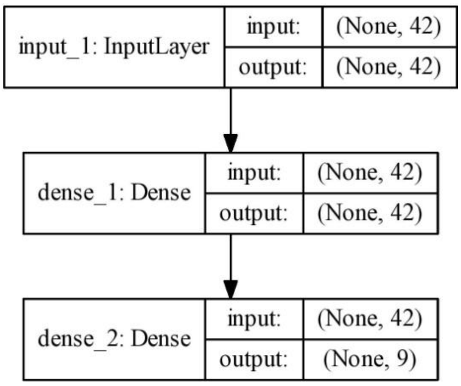

Figure 1: Architecture for student cross subject academic prediction as described in section 2.3

	subject0	subject1	subject2	subject3	subject4	subject5	subject6	subject7	subject8	subject9	...
0	0.8467	0.815	0.83	0.835	0.835	0	0	1	0.6	0.6	...
1	0.7741	0.815	0.66	0.705	0.665	0	0	1	0.7	0.7	...
2	0.82	0.77	0.745	0.745	0.845	0	0	1	0.85	0.7	...
3	0.8066	0.9	0.615	0.805	0.835	0	0	1	0.7	0.6	...
4	0.733	0.735	0.7	0.735	0.585	0	0	1	0.7	0.5	...

Figure 2: Input data representation for neural network

2.3.1InputRepresentation

The data is first cleaned of any absentees and all the grades are converted into grade points as per university norms.This is then divided by 10 to rescale it to [0,1] range. This kind of representation helps the neural net- work converge quickly.The input data shape is (number of students,number of subjects).Each represented a par- ticular student and columns their grade points.

2.3.2Model Architecture

The core learning part occurs in the hidden layer(givenin fig. 1).This layer learns the importance of individual sub- jects to output subjects and how they are related grade- wise this embeds the input marks in to much dense representation

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 42)                0
_________________________________________________________________
dense_1 (Dense)              (None, 42)                1806
_________________________________________________________________
dense_2 (Dense)              (None, 9)                 387
=================================================================
Total params: 2,193
Trainable params: 2,193
Non-trainable params: 0
_________________________________________________________________
```

Figure 3: Trainable parameters on different layers

which is then used by the output layer con- taining 9 neurons as we have 9 different subjects for the semester we are trying to predict.This architecture pro- duces the grade points a particular student might get for the semester the network has been trained for.The op- timizer we used is Adam with a learning rate = 0.001 and activation function is tanh. This combination proved to be the best out of multiple combinations we tried us- ingrelu,sigmoid and tanh.We also found that Nestroven- abled Adam-Nadam is also a good optimizer but we chose Adam as we got better results on test set. Mean Squared Error was proved to be the best loss function for this use case.

$$L(Y_i, \hat{Y}_i) = (1/n) * \sum_{i=1} (Y_i - \hat{Y}_i)^2$$

Here, Yi is the true value,Yˆ I is the predictions of model and n being the number of samples used for evaluation. The complete architecture is illustrated in figure 1

Retrieval Number: K125409811S19/2019©BEIESP
DOI: 10.35940/ijitee.K1254.09811S19

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

III. DATA AND TRAINING

3.1 Kerala Technical University (KTU) Data

We collected our data from the PRS College of Engineering affiliated to Kerala Technical University.Wetook scores of previous 5 semesters and the student's performance in plus two which was preprocessed and cleaned of any abnormalities before feeding to network. Wemade sure the data didn't cont ain any absentees or null values and a particular student was present throughout all the semesters in the same order as he/she appeared in the previous semesters and all semesters were concatenated into one numpyarray, which was fed into the networkfor training. Similary the subjects of last semester wasar- ranged in order and scaled down and fed aslabels. The whole dataset was divided into 90 percent and 10
The whole dataset was divided into 90 percent and 10 percent for training and testing setres pectively. This made sure both training data and testing data came from the same distribution.

3.2 Training

Both baseline and our model was trained using the same training data and was validated on 50 percent of test data. Basel in emodel was trained with default parameter sasdescribed in it'sresearch paper and our model was trained for 3000 epochs with a batch size of 64 and data was shuffled after each epoch. The weight soft he model which showed lowest root mean squared error on validation test was saved.

IV. PREDICTION PERFORMANCE

a. Evaluation Measure

To demonstrate the actual power ofthismodelwefocused on the closeness of grade point prediction to actual grade point that student got.The dataset happened to be imbalanced, considering the grade classification of students.In our case accuracy is not a accurate metric as some subjects are dominated by one particular grade.

We used Root Mean Squared Error (RMSE) on both scaled and non scaled output of the models to evaluate it. The Scaled RMSE showed how much the actual grade point can vary from the predictedvary. Lower the RMSE, Better the model is.

4.2 Results

Table 1: Result Comparisons

Model	ScaledRMSE	ErrorRange
Baseline	1.5185	2
Deep Regressor	1.0657	1

As we have proved in the above table the baseline model had prediction error in the range of 2 grade points which leads to inaccurate predictions but our simple yet powerful model was able to bring down the error range to 1.This reduction is error range by one unit have major advantages when converting back to grades where our model was able to predict with maximum error offset of only 1grade.

V. CONCLUSION

In this paper, we have successfully applied deep learning to the challenging cross subject student performance prediction problem which so far has not been exploited.Wealso demonstrated the higher capacity of Deep Regres- sor model over gradient boosting algorithms like XG- Boost. Also instead of directly using existing classification methods we have done classification viaregression model to predict the grades of student. This method proved to be much more efficient and accurate.

Four novel properties of our model is(1) it doesn't need any complex preprocessing stepo the rthandividing by 10 (2)it can operate to get predictions even when the dataset is small and imbalanced.(3) it doesn't capture trends due to imbalance like baseline models (4) The cross subject prediction capability with minimum error range which is not shown by any other algorithm.

As future work, incorporating more data from different colleges will again further bring down the scaled error range from 1 to much lower values and improve the deep regressor's impressive performance.

REFERENCES

1) S. Kotsiants. C Pierrakeas and P Pintelas, Prevent- ingstudentdropoutindistancelearningsystemusing machine learning techniques, AI Techniques in Web based Educational system at seventh International ConferenceonKnowledge-basedIntelligentinforma- tion and Engineering systems, pages 3-5, September 2003.
2) Lykourentzou, Giannoukos, Nikolopoulosand Loumosin, Dropout prediction in e-learningcourses through the combination of Machine learning tech- niques, Computers and education, Vol. 53 Number 3, pages 950-965, November 2009.
3) L.G.Moseleyand.D.M Mead, Predicting who will drop out of nursing courses: A Machine learn- ing exercise Nurse Education Today,Vol. 28,Number 4,pages 469-475 May2008
4) Byung-HakKim,EthanViziteiandVarunGanapathi, GritNet: Student Performance Prediction with Deep Learning, August2018.
5) F.Miand D-Y Yeung, Temporal models for predict- ing student dropout in massive open online courses, In proceedings of 15th IEEE International Confer- ence on Data Miming Workshp(ICDMW 2015), pages 256-253,2015.
6) Tianqi Chen and Carlos Guestrin, XGBoost: A Scal- able Tree Boosting System, 2016.
7) DiederikP.Kingma,Jimmy Lei Ba, ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION, Published as a conference paper at ICLR,2015.
8) Timothy Dozat, Incorporating Nesterov Momentum intoAdam.