

MOOCVERSITY - Deep Learning Based Dropout Prediction in MOOCs over Weeks

Vignesh Muthukumar,
Department of Information Technology,
S.S.N. College of Engineering,
Anna University,
Chennai, India
vigneshm15123@it.ssn.edu.in

Dr. N. Bhalaji,
Associate Professor
Department of Information Technology,
S.S.N. College of Engineering,
Anna University,
Chennai, India
bhalajin@ssn.edu.in

Abstract: Massive Open Online Courses (MOOCs) has seen a dramatic increase of participants in the last few years with an exponential growth of internet users all around the world. MOOC allows users to attend lectures of top professors from world class universities. Despite flexible accessibility, the common trend observed in each course is that the number of active participants appears to decrease exponentially as the week's progress. The structure and nature of the courses affects the number of active participants directly. A comprehensive review of the available literature shows that very little intensive work was done using the pattern of user interaction with courses in the field of MOOC data analysis. In this paper, we take an initial step to use the deep learning algorithm to construct the dropout prediction model and produce the predicted individual student dropout probability. Additional improvements are made to optimize the performance of the dropout prediction model and provide the course providers with appropriate interventions based on a temporal prediction mechanism. Our Exploratory Data Analysis demonstrates that there is a strong correlation between click stream actions and successful learner outcomes. Among other features, the deep learning algorithm takes the weekly history of student data into account and thus is able to notice changes in student behaviour over time.

Keywords: Deep Learning, Dropout Prediction, MOOCs, Multilayer Feedforward Neural Network, Clickstream data analysis, Personalized student performance, Student Attrition, Tomek Links.

1 Introduction

As an extension of online learning technologies, MOOCs' rapid development has opened up a new era of education by broadening educational boundaries over the Internet. Massively Open Online Courses (MOOCs) offer tremendous learning opportunities in a scalable form to students. Coursera, HarvardX, and Udacity are some examples of MOOCs. The learners are given privileges in popular MOOC platforms to access video lectures, annotated video slides, transcripts, take up quiz and assignment, complete a grading examination and post their queries in discussion forums. MOOCs typically have no enrollment limit because of their online and open nature and allow anyone to participate or drop out without penalty. However, since this student population is significantly different in terms of degree of

interest, background, and other characteristics, a MOOC offers relatively few opportunities for evaluation and personalization.

As a result, students enrolled in MOOCs, unlike campus-confined students, are much more likely to drop out from the course, which puts the efficacy of the learning technology into question. To gain insight into this majority of participants' learning behaviors, personalized user background information is coupled with the clickstream data of the MOOC platforms to perform significant analysis. Discontinuity in participating in the course becomes the primary motive for our experiment. The experiment uses micro-level course data to examine the pattern of enrolment and participation among students in the Introduction to Programming course offered conjointly by Harvard and MIT. This indicates how the design features can be important for persistence and how the contents are accessed by each student.

The underlying analysis will provide a detailed review of student usage patterns: registration, attendance of lectures, completion of assignments, performed course interaction activities, and obtaining a certificate of completion.

2 Existing System

There are a few existing systems that make use of the user's engagement features with the course and recommend similar courses for them. Different approaches have been designed with the aim of understanding and analysing learner interaction in MOOCs efficiently. The authors in [1] identify four different learner engagement classes within MOOCs based on two core attributes: accessing video lectures and grades obtained from assignments. These classes are Completing, Auditing, Disengagement, and Sampling. In this case, the authors used clustering techniques to describe engagement activity in MOOCs.

The research conducted in [2] highlights the prediction of dropout using the activity features of the learner. The dropout was predicted based on the students' total absence time instead of the student's last visit to the course. However, the basic outline for building this thesis emerged from the extensive work undertaken in [3]. By considering the users' clickstream data, the authors presented a novel solution to predict weekly dropout rates in a particular course. They provided a comparative study between different machine learning algorithms and decided that the best accuracy was provided by the algorithm Random Forest and Support Vector Machines. This experiment was performed on fixed-term dropout data. This provided an extensive scope for improvising the model using Deep Learning algorithms by using temporal data as they tend to deliver more sophisticated outcomes.

In [4], Feed forward Neural Network has been implemented to predict student attention in MOOCs, taking into account student sentiments. In this case, only the behavioural attributes are used to measure the performance of learners. Research in [5] includes building the model of fixed - term dropout prediction that can identify all students at risk at once. While the clickstream data was used to

understand the student's interaction with the course, [7] gave a breakthrough idea that helped the designers of MOOCs identify which part of the videos might need editing. The model uses clickstream information to develop methodology for finding prediction of next click state, in - video, and complete prediction of course dropout. [8] Discusses how Neural Networks reinforces the sentiment analysis process that predicts the attrition of students in MOOCs. This used discussion forums, threads and votes obtained by each post to conduct the analysis.

2.1 Drawbacks of this System

As mentioned earlier, the difference in the proportion of enrolment and completion of a course is very drastic and this problem has to be addressed. The earlier systems do not consider the personalized information about the users as each person enrolling for a course would have different motives and would be from a different background. Failing to conceptualize this, the course providers will not be aware about the reasons for dropout. These systems talk about the model of fixed - term dropout prediction that can identify all students at risk at once. However, given the large number of dropout students in MOOCs, instructors are unable to provide effective feedback at the same time to many at - risk students.

3 Proposed System

The main goal of **MOOCiversity** is to predict student dropout in a particular MOOC over weeks. We consider the Introduction to Programming MOOC jointly launched by Harvard and MIT for our experiment.

The system is implemented using multilayer feed forward neural network architecture. A multilayer feed forward neural network is an interconnection of perceptron in which data and calculations flow from input data to outputs in a single direction. It is fed weekly information from all the students who have enrolled 5 weeks for the course. This includes personal user details as well as clickstream details. Click stream data involves data like number of times a video is played, number of attempts taken on a graded quiz, etc.

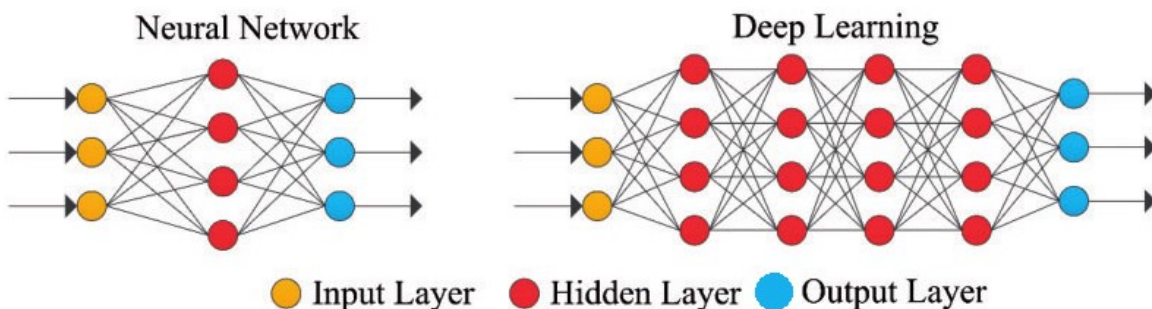


Fig 3.1 (Multilayer layer feed forward neural network)

The proposed system makes use of these data points to accurately predict two things: number of students expected to drop out in any given week and probability of a student to drop out in any given week.

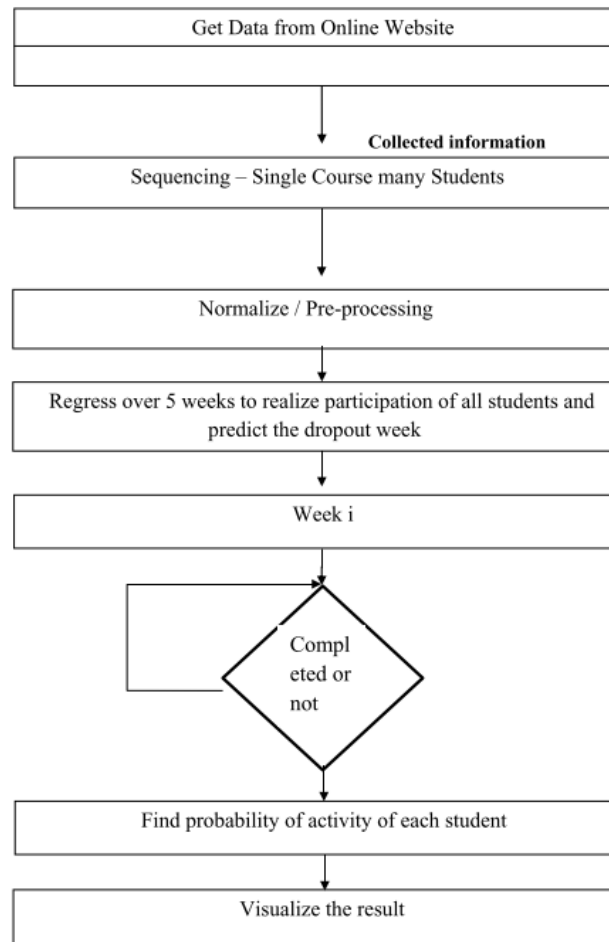


Fig 3.2 (Workflow of proposed model)

The system also provides a comparative study between various models to find which of them performs accurately. SVM, Naive Bayes, Logistic Regression, Linear Discriminant Analysis and Neural Networks are some of the models used. We use metrics such as Area under the curve, Precision, Recall, and Confusion Matrix to compare the results.

3.1 Methodology

Before selecting the best model to create our system, different algorithms were tested to compare their effectiveness in producing the result. We have essentially trained machine learning algorithms to narrow the gap between actual and predicted values. For this purpose, as our response variable is nominal, we split the data into (85-15) ratio with stratified sampling.

3.1.1 Data Preprocessing

The data used in this study consists of 5 log files each with 83383 data entities representing the activities of the learners enrolled for Introduction to Programming MOOC, where each row is a single user session. Because our dataset consisted of imbalanced class labels, intensive preprocessing has been done.

The **K - Nearest Neighbor** algorithm had been used to estimate the missing values by selecting neighbouring values. The major reason for choosing KNN is that it can be used for data that are continuous, discrete, ordinal and categorical. The next kind of preprocessing was converting the categorical values into numerical values. This was done using **One-Hot Encoding**. Since the data was highly imbalanced, the minority classes had to be resampled to create a balance. Once adequate quantities of records were available for minority class, the system performed majority class undersampling using a method called **Tomek Links**. Finally **Correlation Matrix** was used to extract the best set of features from the entire dataset.

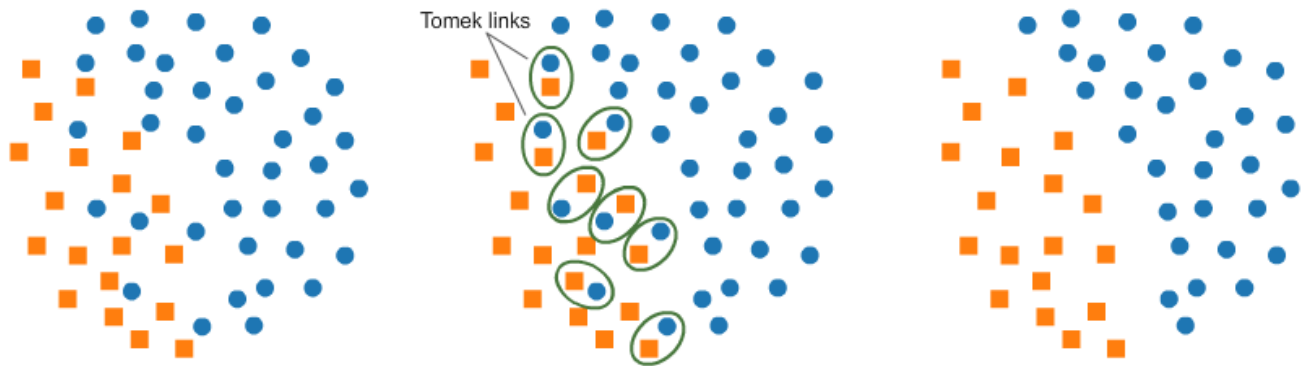


Fig 3.3 (Tomek Links)

Once the dataset was suitable for model creation, we used different methodologies for selecting features to get the most important ones. Some of the machine learning models are Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT). Along with this a Multilayer feed forward (MLF) neural network model was also used. Among the six models, the neural network showed better precision and recall values and was chosen to build the entire system.

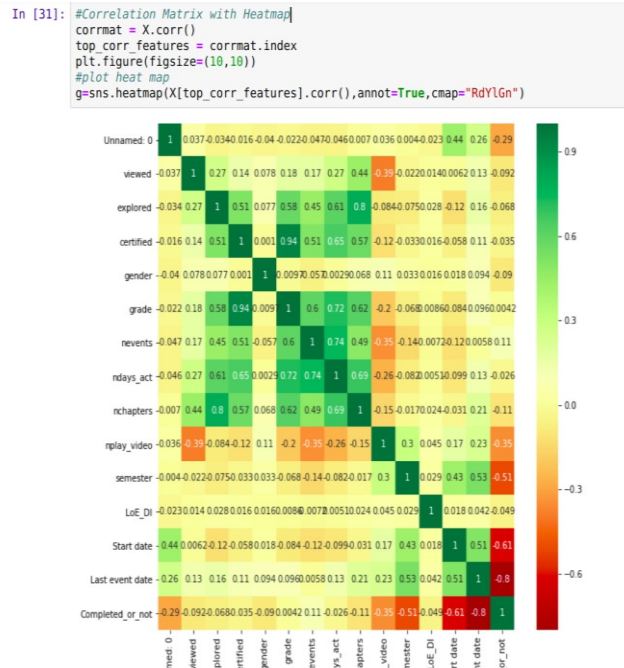


Fig 3.4 (Correlation Matrix)

3.1.2 Multilayer Feedforward Neural Network Algorithm

Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. A good way to think of an NN is as a **composite function**. You give it some input and it gives you some output. There are three parts that make up the architecture of a basic NN. The model was simulated using the following parameter settings to obtain a very efficient system:

- **Number of input features:** 9
- **Number of hidden layers:** 4
- **Activation functions:** Softmax, ReLU, Sigmoid
- **Number of epochs:**150
- **Batch size:** 35
- **Validation split:** 85% train and 15% validation
- **Class weight:** auto
- **Loss function:** Binary cross entropy
- **Optimizer:** Adam
- **Metrics:** Accuracy, Precision, Recall and ROC

$$Z_1 = \sum_{i=1}^3 W_{1,i} x_i - \mu_1 \quad \text{and} \quad Z_2 = \sum_{i=1}^3 W_{2,i} x_i - \mu_2$$

$$Y_1 = g_1(Z_1) = g_1\left(\sum_{i=1}^3 W_{1,i} x_i - \mu_1\right)$$

and

$$Y_2 = g_2(Z_2) = g_2\left(\sum_{i=1}^3 W_{2,i} x_i - \mu_2\right)$$

4 Observations and Results

A total of 6 different models were experimented and the prediction probability were compared using various metrics.

Models	Precision score for Class 1	Recall Score for Class 1
Support Vector Machine	78	73
Logistic Regression	57	94
Naive Bayes	49	96
Linear Discriminant Analysis	67	92
Decision Tree	63	92
Neural Network	72	84

Table 4.1 (Performance comparison of various models)

From the above table it can be clearly seen that Neural networks have the best precision metrics amounting to upto 72. Linear discriminant analysis also has a good precision metrics amounting to 67. Other models like Decision tree, Naive Bayes and Logistic regression have fairly low precision metrics. Hence the model was created using neural networks and the weekly dropout probability is represented in a distributed format.

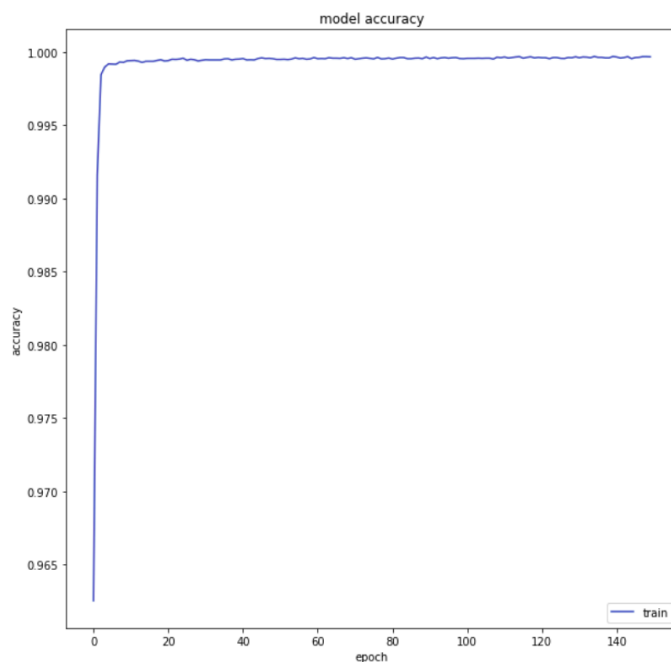


Fig 4.1 (Model Accuracy)

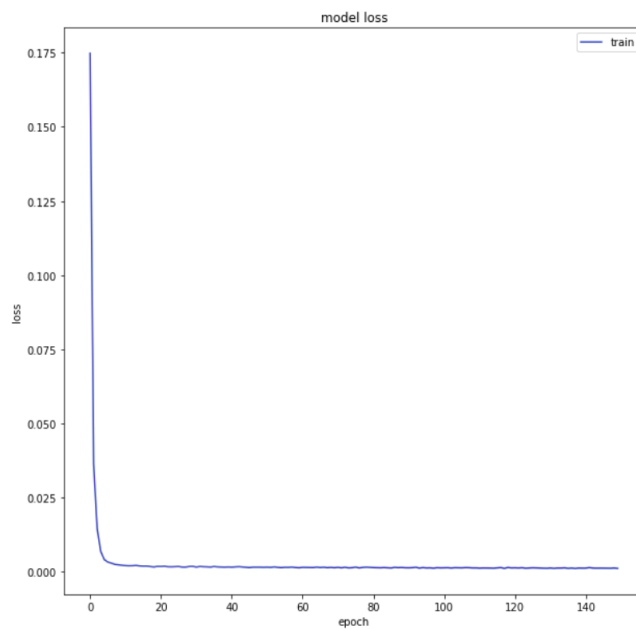


Fig 4.2 (Model Loss)

	precision	recall	f1-score	support
0	1.00	0.99	0.99	11501
1	0.72	0.84	0.78	354
micro avg	0.99	0.99	0.99	11855
macro avg	0.86	0.92	0.88	11855
weighted avg	0.99	0.99	0.99	11855

Recall score: 0.8418079096045198
 Precision score: 0.9868324748176166

Fig 4.3 (Precision and Recall of MLF NN)

Confusion matrix, without normalization
 [[11384 117]
 [56 298]]

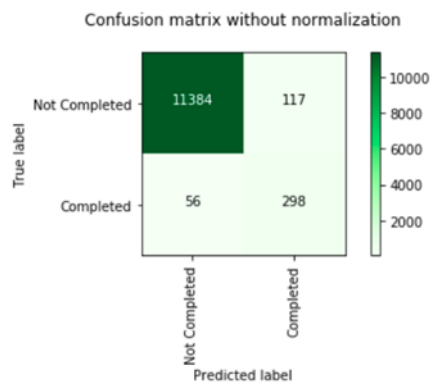


Fig 4.4 (Confusion Matrix of MLF NN)

After testing out the system for 5 weeks of the user's dataset, the system proved to determine the number of students who would dropout over the weeks. A binary class output is provided by the model: learners who have completed and those who have not.

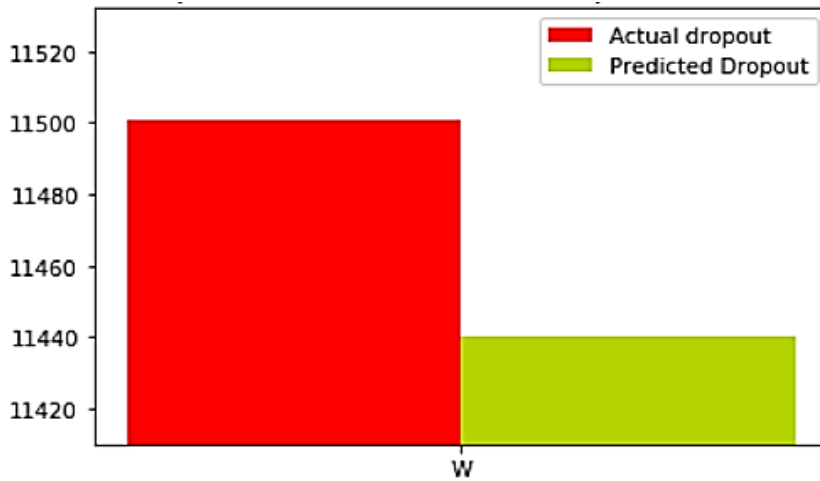


Fig 4.5 (Actual vs Predicted Dropout in a Week)

This model also provides the probability with which learners complete the week. Since the performance of the learner over weeks is interrelated, the model considers the history of the probability that the previous weeks were completed in order to provide the classification of the following weeks.

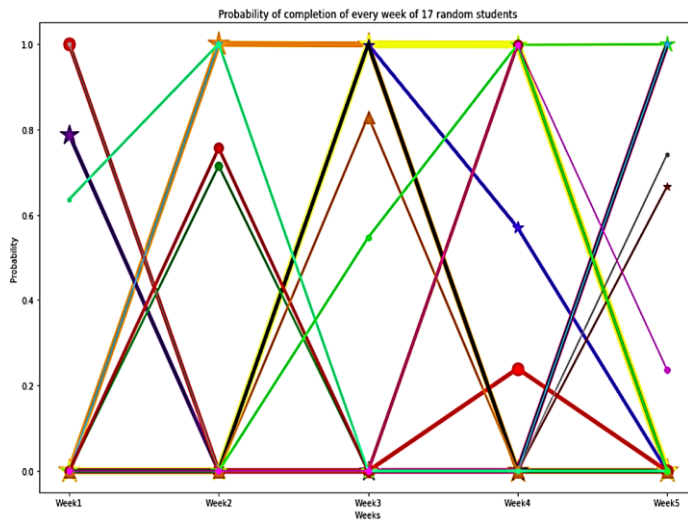


Fig 4.6 (Students' probability of completion over weeks)

Finally, the course instructors are given a detailed weekly analysis where they can find out which week has the most dropouts and why.

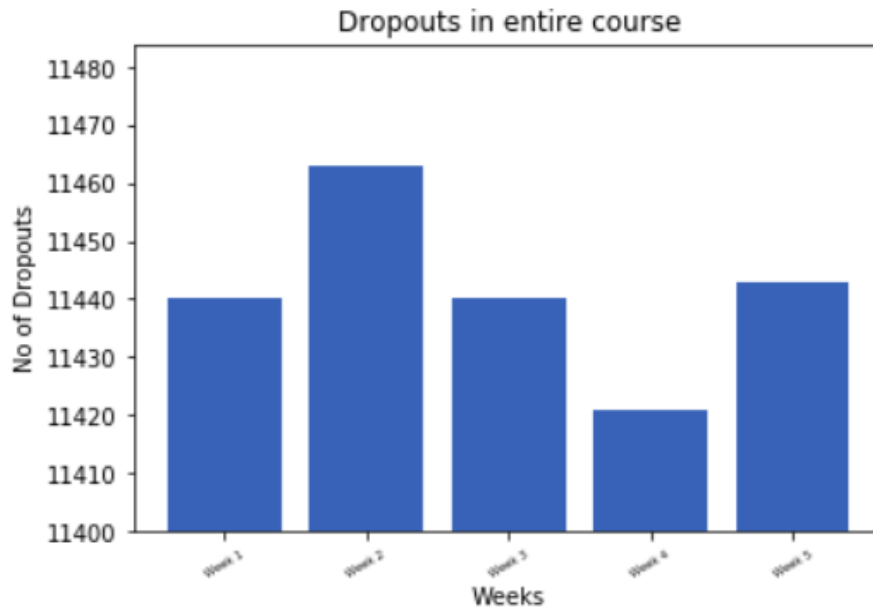


Fig 4.7 (Week-wise dropout)

Each student also obtains a weekly rank for their performance. A continuous dip in the rank would alert the learner about his/her likelihood of dropping the course in the upcoming weeks, while a steady increase will motivate the student to complete the course early.

Week: 1 - Rank: 11306

Buck up buddy

Week: 2 - Rank: 918

That was a great improvement

Week: 3 - Rank: 6036

Watchout! You are slipping

Week: 4 - Rank: 11419

Watchout! You are slipping

Week: 5 - Rank: 11454

Watchout! You are slipping

At this rate you are more likely to dropout

Fig 4.8 (Intervention for a student after every week)

5 Further Work

There is vast scope for further work. Currently, the personalization and prioritization only show statistical validity. Researchers can further examine their validity by implementing the proposed model of prediction and intervention personalization method in actual MOOC courses to evaluate whether this approach can reduce the dropout of students.

Similar experiments on more general online courses can be carried out to examine whether deep learning can be useful in other educational contexts. Future studies can explore methods to further improve the performance of deep learning predictions in MOOCs, such as increasing the network's hidden layers. Also a complete sentiment analysis can be performed with the posts made by the users in the forums. More efforts can be taken to obtain additional information about clickstream features so that content wise interventions can also be given. For example, the video segment where most of the students skipped means something is wrong with the video and has to be changed.

6 Conclusion

As seen and observed above the proposed system employing Deep Neural Networks produces a marked improvement to the performance of the existing and default mode of dropout prediction. The deep learning approach will not only build more accurate dropout prediction models compared to baseline models, but will also introduce a valid approach to inform intervention design, personalizing and prioritizing support for students at risk using MOOC dropout probabilities.

Significant variance in data was an important factor as to why some predictions were out of class. Another important factor was the inability to reflect the mental thinking process of the student regarding the content and materials of the course. Conclusively, the dropout prediction model that was constructed using the deep learning algorithm proved to be way more effective and efficient.

REFERENCES

- [1] A. D. Ho et al., "HarvardX and MITx: The First Year of Open Online Courses, Fall 2012-Summer 2013," SSRN Electron. JNo. 1, pp. 1–33, 2014.
- [2] Abidi, S.M.R.; Hussain, M.; Xu, Y.; Zhang, W. "Prediction of Confusion Attempting Algebra Homework in an Intelligent Tutoring System through Machine Learning Techniques for Educational Sustainable Development." Sustainability 2019, 11, 105.
- [3] Al-Shabandar, R., Hussain, A, Laws, A., Keight, R., & Lunn, J. (2017). "Machine learning approaches to predict learning outcomes in massive open online courses." Paper presented at the 2017 international joint conference on neural networks (IJCNN 2017), Anchorage, Alaska, USA.
- [4] C. Peng, K. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," J. Educ. Res., vol. 96, pp. 3–14, 2002.

- [5] D. S. Chaplot, E. Rhim, and J. Kim, “Predicting student attrition in MOOCs using sentiment analysis and neural networks,” *Work. 17th Int. Conf. Artif. Intell. Educ. AIED-WS 2015*, vol. 1432, pp. 7–12, 2015.
- [6] D. S. Chaplot, E. Rhim, and J. Kim, “Predicting student attrition in MOOCs using sentiment analysis and neural networks,” in *Fourth Workshop on Intelligent Support for Learning in Groups*, Madrid, Spain, June 2015.
- [7] Halawa, S.; Greene, D.; and Mitchell, J. 2014. “Dropout prediction in MOOCs using learner activity features.” In *Proceedings of the European MOOC Summit*.
- [8] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, “Predicting MOOC dropout over weeks using machine learning methods,” in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
- [9] Mongkhonavit, K., Kanopa, K. and Lang, D. (2019). “Deep Knowledge Tracing and Engagement with MOOCs.” *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. Pp. 340-342, Tempe, AZ, USA — March 04 - 08, 2019.
- [10] P. Vizslay, M. Lojka and J. Juhár, “Class-dependent two dimensional linear discriminant analysis using two-pass recognition strategy”, in: *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, IEEE, 2014, pp. 1796– 1800.
- [11] R. F. Kizilcec, C. Piech, and E. Schneider, “Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses,” *Lak '13*, p. 10, 2013.
- [12] R. R. Kabra and R. S. Bichkar, “Performance prediction of engineering students using decision trees,” *Int. J. Comput. Appl.*, vol. 36, no. 11, pp. 8–12, 2011.
- [13] Sinha, T., Jermann, P., Li, N., Dillenbourg, P. (2014). “Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions.” *Proc. of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*.
- [14] Svozil D, Kvasnicka V, Pospichal J. “Introduction to multi-layer feed-forward neural networks.” *Chemometrics and intelligent laboratory systems* 1997;39(1):43-62.
- [15] Tharwat, Alaa & Gaber, Tarek & Ibrahim, Abdelhameed & Hassanien, Aboul Ella. (2017). “Linear discriminant analysis: A detailed tutorial.” *Ai Communications*. 30. 169-190., 10.3233/AIC-170729.
- [16] W. Xing and D. Du, “Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention,” *J. Educ. Comput. Res.*, 2018.
- [17] Scipy. Retrieved from <https://docs.scipy.org/doc/scipy/reference/tutorial/general.html>