

# SAP: Student Attrition Predictor

Devendra Singh Chaplot  
Samsung Electronics Co., Ltd.  
Seoul, South Korea  
dev.chaplot@samsung.com

Eunhee Rhim  
Samsung Electronics Co., Ltd.  
Seoul, South Korea  
eunhee.rhim@samsung.com

Jihie Kim  
Samsung Electronics Co., Ltd.  
Seoul, South Korea  
jihie.kim@samsung.com

## ABSTRACT

Increasing rates of student drop-outs with increase in popularity of Massive Open Online Courses (MOOCs) makes predicting student attrition an important problem to solve. Recently, we developed an algorithm based on artificial neural network for predicting student attrition in MOOCs using student sentiments. In this paper, we present a web-based tool based on our algorithm which can be used by educators to predict and reduce attrition during a course and by researchers to design and train their own system to predict student attrition.

## Keywords

Student Attrition, MOOC, Sentiment Analysis, Neural Network, Educational Data Mining, Student Drop-out

## 1. OVERVIEW

Growing popularity of MOOCs is attributed to their accessibility, scalability and flexibility. With scalability, MOOCs also provide huge amounts of data of student activity which can be used to predict their behavior. We have developed an algorithm to predict student attrition [4] which uses click-stream log and forum posts from MOOCs to extract features such as number of page views, clicks, study sessions, etc. as suggested by previous studies [1, 3, 5, 6]. A unique feature used by our algorithm is student sentiments in forum posts, which is calculated using lexicon-based Sentiment Analysis with SentiWordNet 3.0 [2] as the knowledge resource. The values of all these features for current week are passed as inputs into an artificial neural network, whose output indicates whether student is going to drop out in the following week. Using data from Coursera course 'Introduction to Psychology', we get 74.4% accuracy with false negative ratio of 0.136, leading to a Cohen's Kappa value of 0.435.

## 2. STUDENT ATTRITION PREDICTOR

We present a web tool having three interfaces for educators and researchers to predict and study student attrition.

## 2.1 Sentiment Analysis

Sentiment Analysis of student's forum posts is the unique feature which wasn't used by previous algorithms and improves the Cohen's Kappa value of our algorithm by about 13%. Effectiveness of using sentiment analysis can be seen by the changes in results from neural network when student sentiments are added as input. Our tool also provides option to get the Sentiment score of any student's forum post.

## 2.2 Pre-trained Neural Network

Users have the option to use our pre-trained neural network to predict student drop-out. This allows our tool to be used freely by educators to predict student attrition. Since we predict whether student is going to drop-out in the following week and not whether student is going to complete the course, our algorithm pin-points the exact week when student is predicted to drop-out and thus, educators can use our tool during the course in order to take necessary student-specific actions to prevent or reduce attrition. Apart from MOOCs, Student Attrition Predictor can also be used by traditional classroom setting educators, using digital mediums for study and interaction in schools, which are becoming increasingly popular in recent years.

## 2.3 Design new Neural Network

Our tool also provides an interactive graphical interface for the users to design their own unique neural network. A screenshot of design interface is shown in Figure 1. It shows an input panel, training and testing data panels, a neural network design canvas and a results panel. The process of using Design interface can be divided into 3 phases:-

- **Design:** Users can add their own nodes in the 'Input' panel and select any number of hidden layer nodes. The canvas in the middle of Figure 1 shows the structure of designed neural network.
- **Train:** Training data can be uploaded in 'Training Data' panel and used to train the designed neural network. Options for selecting number of training iterations, classification boundary and learning heuristic (like back-propagation, resilient propagation, etc.) for training Neural Network will also be provided.
- **Test:** After training, individual input values can be entered in the input panel or test data can be uploaded in 'Test Data' panel to get results from trained neural network. 'Results' panel shows metrics such as Accuracy, False Negative Rate and Cohen's Kappa value.

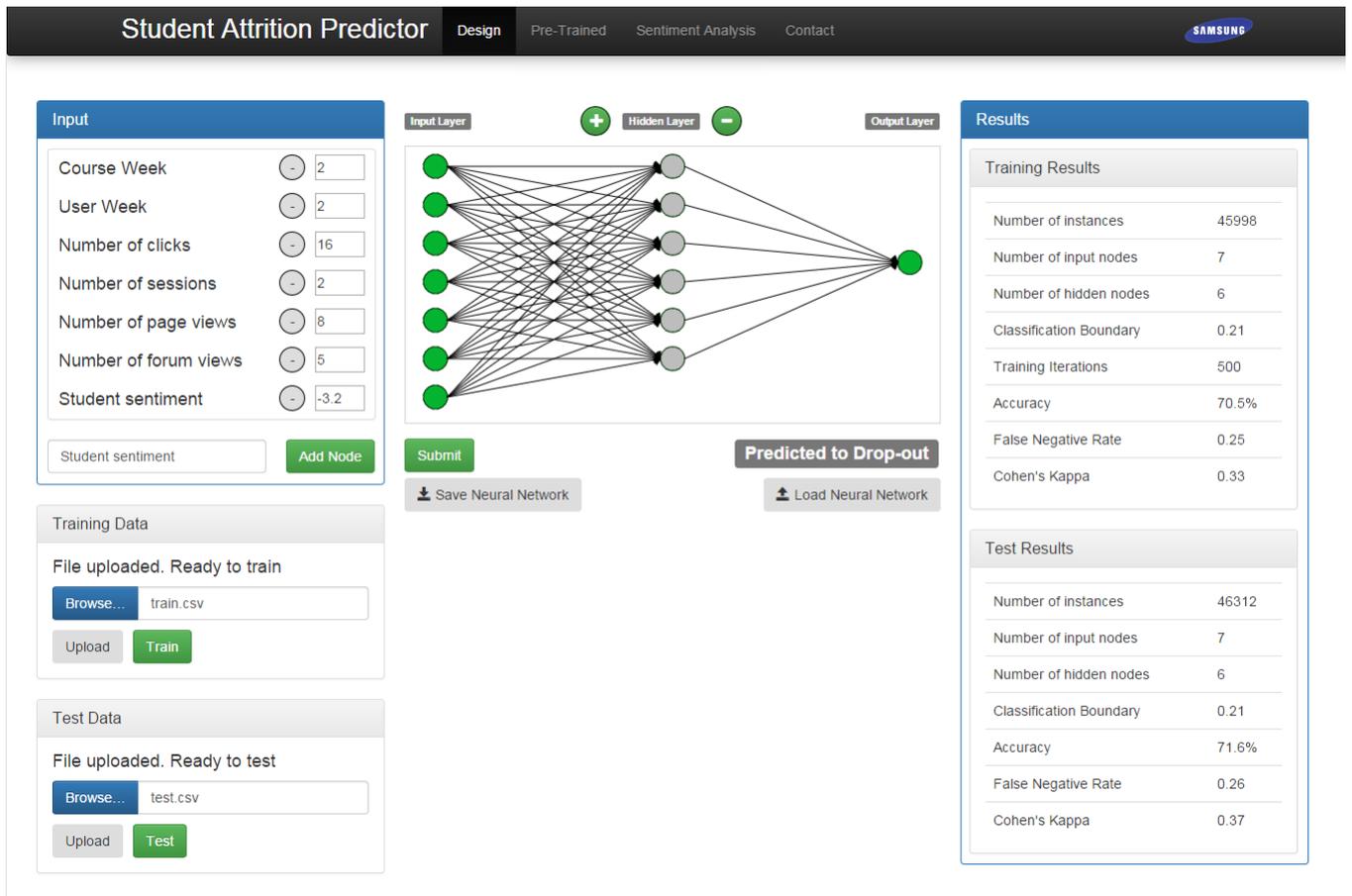


Figure 1: Screenshot of Design Interface of Student Attrition Predictor

This interface is especially useful for researchers who can decide the input features and structure of their own neural network, train and test it by uploading their own data and optimize the parameters and learning heuristic according to their application. The designed and trained neural network can be saved and loaded into the tool at any point.

### 3. CONCLUSION

There has been lot of research in recent years on predicting student attrition. In contrast to many studies trying to find reasons behind attrition, we focus on predicting and reducing attrition. Student Attrition Predictor not only predicts student drop-out, but also identifies the precise week when student is likely to drop-out in order to reduce attrition during the course. To the best of our knowledge, there is no direct way for educators to benefit from years of research on predicting student attrition. This tool acts as a medium for educators to directly utilize our research in this field. The tool also provides an easy graphical interface to researchers for further experiments.

### 4. REFERENCES

- [1] B. Amnueypornsakul, S. Bhat, and P. Chinprutthiwong. Predicting Attrition Along the Way: The UIUC Model. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social*
- [2] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010.
- [3] G. Balakrishnan. Predicting student retention in massive open online courses using hidden markov models. Master's thesis, EECS Department, University of California, Berkeley, May 2013.
- [4] D. S. Chaplot, E. Rhim, and J. Kim. Predicting student attrition in moocs using sentiment analysis and neural networks. In *Proceedings of AIED 2015 Fourth Workshop on Intelligent Support for Learning in Groups*, Madrid, Spain, 2015.
- [5] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart. Predicting MOOC Dropout over Weeks Using Machine Learning Methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65, Doha, Qatar, October 2014.
- [6] M. Sharkey and R. Sanders. A Process for Predicting MOOC Attrition. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 50–54, Doha, Qatar, October 2014.

*Interaction in MOOCs*, pages 55–59, Doha, Qatar, October 2014.