

# Predicting Student's Academic Performance in a MOOC Environment

Rahila Umer<sup>1</sup>, Teo Susnjak<sup>1</sup>, Anuradha Mathrani<sup>1</sup> and Suriadi Suriadi<sup>2</sup>

<sup>1</sup>Institute of Natural and Mathematical Science, Massey University, Auckland, New Zealand.

<sup>2</sup>Queensland University of Technology, Brisbane.

(r.umer@massey.ac.nz, T.Susnjak@massey.ac.nz, A.S.Mathrani@massey.ac.nz, s.suriadi@qut.edu.au)

**Abstract:** Massive open online courses (MOOCs) provide an opportunity for students to register for courses offered by best universities around the world. There are massive enrollments in the courses offered in MOOCs; however they suffer from low completion rate and low student retention. To address these problems, it is necessary to make early prediction of students' academic performance to enable targeted and timely interventions for those students at risk of non-completion. This study proposes the use of machine learning algorithms to predict students' academic performance in a MOOCs environment and compared the predictive power of four machine learning algorithms: Logistic Regression, Naïve Bayes, Random Forest and K Nearest Neighbor. Our results show that performance of predictive models are promising and can be used for early prediction of students that are likely to fail in a course. Furthermore, Random forest classifier outperforms other classifiers in a statistical significant way.

**Keywords:** Learning analytics, MOOCs, Machine learning, Data mining, Prediction

## 1. Introduction

Massive open on-line courses (MOOCs) provide an opportunity for students to register for courses offered by best universities around the world. Students enrolled in the courses have diverse goals and motivations. Anyone with access to the Internet can register for any high quality and advanced courses offered. MOOCs comprise of video lectures, quizzes, reading resources and provide a forum platform where students can do productive discussion and engage with peers anytime and from anywhere.

All activities of student's are recorded in the MOOCs environment, opening up an opportunity to analyse the large amount of data to gain evidence-based understanding of the behaviour of students in such an environment. Despite the large number of enrolments, student retention in MOOCs is low, often less than 20% [1] and is therefore heavily criticized. In order to increase the retention rate, one could exploit the massive amount of data to predict the likelihood of dropout or failure of the course, thus enabling effective early intervention strategy for those students who are struggling during the course by offering relevant and targeted help.

Learning Analytics (LA) and Educational Data Mining (EDM) are two approaches that focus on data-driven techniques to inform research on practices to mitigate issues like drop-outs, failure rates and low retention rates. Students' digital traces which they leave behind as a result of their online interactions, such as clicks, page visited, and video watched, are recorded in log history during the course [2]. Campbell and Oblinger [3] presented a 5-step method (including captures, report, predict, act and refine) as the main theme in LA. Once, data related to students' interactions with a course is captured and reported, they can be analysed to make some predictions about students performances, informing the design of the subsequent pedagogical interventions (act).









