MOOC *next week* dropout prediction: weekly assessing time and learning patterns

Ahmed Alamri[1], Zhongtian Sun[1], Alexandra I. Cristea[1], Craig Stewart[1], Filipe Dwan Pereira[2]

[1]Department of Computer Science Durham University, [2] Institute of Computing Federal University of Roraima
[ahmed.s.alamri, alexandra.i.cristea]@durham.ac.uk

Abstract. Although Massive Open Online Course (MOOC) systems have become more prevalent in recent years, associated student attrition rates are still a major drawback. In the past decade, many researchers have sought to explore the reasons behind learner attrition or lack of interest. A growing body of literature recognises the importance of the early prediction of student attrition from MOOCs, since it can lead to timely interventions. Among them, most are concerned with identifying the best features for the entire course dropout prediction. This study focuses on innovations in predicting student dropout rates by examining their *next-week-based learning activities and behaviours*. The study is based on multiple MOOC platforms including 251,662 students from 7 courses with 29 runs spanning in 2013 to 2018. This study aims to build a generalised early predictive model for the weekly prediction of student completion using machine learning algorithms. In addition, this study is the first to use a '*learner's jumping behaviour*' as a feature, to obtain a high dropout prediction accuracy.

Keywords: Learning Analytics, Early Dropout Prediction, Machine Learning, Behavioural Pattern

1 Introduction

Massive Open Online Courses (MOOCs) offer open access courses to unlimited learners in an online learning manner. MOOC as a term was first coined in 2008, followed by the naming of 2012 as the 'Year of the MOOC', when MOOC providers, such as Coursera, Udacity, edX and FutureLearn, were all launched and they have reached to millions of learners across the world [26, 14], MOOCs have proven a popular education choice and become a critical mainstream approach to democratise knowledge [12]. However, it should be noticed that only 3-15% of participants complete their courses [5]. Such a situation undermines the initial purpose of MOOC that provides free access for massive numbers of students. Therefore, academics are interested in exploring why participants drop out and how to improve their engagement with the course until completion [2].

Researchers intend to find the most predictive feature(s) of students' dropout activity and thus enable early intervention. One usual way is to identify learning behaviour indicators to raise precision and recall of MOOCs' completion prediction [5]. However, data is not always available for such indicator analysis. For instance, non-completion can be predicted by a linguistic analysis of discussion forum data [24]. Nevertheless, as students' comments only amount to 5-10% of posts in discussion forums, this feature is not applicable universally [21]. Additionally, numerous variables can be considered for non-completion analysis, such as student profile data (e.g., country, age, gender) and course-attended related data (e.g., reading, watching, writing, taking quizzes). To the best of the authors' knowledge, this study is the first to consider participants' learning paths and associated behaviours in weekly dropout prediction. According to [1], a learning path is an insightful dropout prediction feature as successful learners will follow the instructed path and exhibit the so-called catch-up learning behaviours. Conversely, learners may jump forward and backward in their learning sessions [7], defined as exhibiting jumping behaviour and they are more likely to quit in the process. This study also considers other features such as number of learning activities, to predict student completion in the following week. Hence, our research question and its respective sub-questions are formulated as follows:

1. Are there (high) differences in the prediction of weekly dropout and whole course dropout?

2. Will the weekly predictive model be more accurate after considering student jumping behaviours and catch-up learning patterns during the course?

The main contributions of this paper are:

- We compare the prediction of *weekly dropout* and dropout of the whole course.
- New feature: we are the first to incorporate students' learning patterns, specifically *jumping behaviours* into the weekly predictive model and demonstrate the effectiveness of it.
- We implement seven machine learning algorithms and demonstrate that our proposed method outperforms the current best-in-class.

2 Related work

Under the context of MOOCs' rapid spread to millions of people, the low completion ratio encourages researchers to explore, reason and build prediction models for dropout since 2014 [8]. The prediction of MOOC completion, especially at an early stage, has been the primary concern of researchers in learning analytics. Existing studies mainly analyse long-term learner behaviours, i.e., discussion activity, clickstream data, and time spent, based on different machine learning (ML) methods. For instance [14] examined learners' study pattern under a predictive ML framework for a 12-week-long psychology MOOC course. They improved 15% in prediction accuracy (70% up to 85%), compared to baseline methods. However, the proposed model did not perform well at early dropout detection.

[25], targeted struggling learners who need early intervention to keep the engagement, by designing a prioritising at-risk student temporal model. They illustrated the necessity of building an effective and robust ensemble stacking prediction model for such analysis. [36], used data from the first two weeks of study, to allow for early intervention and they achieved accuracy of 80%.

Another study, [10], generated an average of 92% precision and 88% accuracy result of dropout prediction based on a two-layer cascading classifier structure. Additionally, [16] built an ML-based sliding window model based on Support Vector Machine for course completion prediction, which allowed MOOC instructors and designers to track potential dropouts.

However, all the above studies mainly focus on predicting participants' dropout activities for the entire course rather than in the upcoming week.

This paper focuses on predicting students' weekly completion, which we define (following the overall completion in other studies [13], applied to the week level) as accomplishing 80% of learning activity in the following week during the entire course. For example, we will predict students' completion of the second week by using their previous learning behaviours in the first week only. In addition, the model will predict students' completion of the fifth week by using their previous four weeks learning pattern. [11] demonstrated that clickstream-based features are much more predictive for drop out study. This paper will mainly use clickstream-based learning topics accessed for prediction. Additionally, according to [1], participants' learning patterns (linear learning behaviour followed by instructed learning path or jumping learning behaviour opposite the former) are an insightful feature for drop out prediction. We are the first to incorporate the students' learning patterns into our weekly drop out predictive model by reviewing their previous behaviours.

3 Methodology

Future Learn is one of the youngest massive online learning platforms (since 2012), and the European counterpart to USA's Coursera, EdX, etc., which now supports 327 courses created by 83 partners and reached 3 million students by 2018 [7]. As it is a newer platform, there are fewer studies performed on it. We fill this gap by selecting courses delivered through it. This study analyses a massively large dataset of 29 runs (Each course has run several times over years) of 7 multidisciplinary courses which falls under four main categories: Computer Science, Literature, Business and Psychology. The courses have been delivered through FutureLearn by two universities in the United Kingdom (University of Warwick and Durham University) between 2013 and 2018. The studied courses have a length of 4 to 10 weeks. The structure of these courses is based on a weekly learning unit. Every learning week includes so-called 'steps', which cover images, videos, artic les and quizzes. Having joined a given course, learners can access these steps and optionally mark them as completed or solved quizzes. These steps also allow comments, replies and likes on these comments, from different users enrolled within the course. Moreover, quizzes can be frequently attempted, until the correct answer is obtained.

We use raw data and aggregate data, i.e., data composed from different raw data sources. We use data for early prediction as well as for general descriptive analysis. We employ data generated with various techniques: e.g., generated applying sentiment analysis on student information exchange, and to limit it somewhat for the current paper, we have decided to perform a first aggregation step based on the weekly learning unit, which is used as a **synchronisation** point in instructor led FutureLearn courses.

In total, we have obtained interactional educational data (not publicly available) for 251,662 students shown as below in Table 1. Enrolled refers to registered students and accessed refers to students who have accessed the course at least once. It can be seen from the data in Table 1 that about half of enrolled students in MOOC do not access the course contents after the course has started. Each course has several runs as they are popular and held for more than one term. 'The Mind Is Flat' is the largest course among others in term of enrolled students, accessed students and number of runs see Table 1.

Table.1 courses' summary

Course	*Enrolled*	*Accessed*	*Run*
Open Innovation in Business (OI)	*6071*	*2798*	*3*
Leading and Managing People-Centred Change (LMPCC)	*10417*	*6575*	*3*
Babies in Mind (BIM)	*48771*	*26175*	*6*
Big Data (BD)	*33427*	*16272*	*3*
Shakespeare (SHK)	*63625*	*29432*	*5*
Supply Chains (SUP)	*5808*	*2912*	*2*
The Mind is Flat (THM)	*83543*	*39894*	*7*
Total	***251662***	***124058***	***29***

As we have used a massive dataset for different courses, we have prepared the training and testing sets based the last Run. For example, in the The Mind Is Flat course, we extracted data from several runs (1-6), with students activities between 2013 to 2017, to train our models, and to test the model, we used a new data set from a different Run (Run 7) that contains students' activities in 2018 - see Figure 1 - which is similar to some extent to transformer models [23].

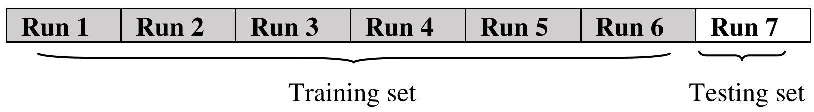

Fig.1 The Mind Is Flat course

Moreover, we have incorporated students' learning patterns, specifically *jumping behaviours*, for now, into the weekly predictive model, by adding a new column that presents number of jumping activates for each student in each week. To demonstrate the effectiveness of jumping behaviours, we compared the performance of weekly prediction models with and without students' jumping behaviours. In addition, we run the features' importance to identify the best indicators (features) to predict student dropout in each week.

3.1 Sentiment Analysis

In this research, the power of Natural Language Processing (NLP) has been used to analyse student comments and use them as features to predict their dropout activities. A tool called Textblob[4] has been employed, in order to classify students' comments into three categories: *positive*, *neutral* and *negative*. TextBlob is an NLP-oriented Python library, which measures polarity and subjectivity of a textual dataset for certain tasks, such as sentiment analysis, classification, part-of-speech tagging, extraction and more complex text processing tasks [20].

3.2 Weekly Prediction

Although a considerable amount of literature has been published on the prediction of MOOC dropout rates, there is no formal definition of student dropout [22]. Researchers in the domain have been using a variety of definitions. In this current research, we have prepared the dataset based on the *weekly prediction technique*, to determine at-risk students at an early stage. It is believed that predicting at-risk students from their previous weeks' activities may improve the model prediction performance. Therefore, in this study, we have implemented seven predictive models, to provide early intervention for learners at-risk in the following week. Each week, we predict the students who do not access 80% of the topics in the coming week, by using previous week/weeks activities as input for our model. The results are generated by seven chosen ML algorithms. We compared our *weekly prediction method* (see equation 1) with the more traditional method of predicting students' dropout from the whole course (the students who do not access 80% of the whole course, see equation 2).

$$Dropout = \sum_{week=2}^{\infty} accessedsteps < \sum_{week=2}^{\infty} totalsteps \; x \; 0.8 \qquad (1)$$

Weekly dropout prediction **(WP)**

$$Dropout = \sum_{allweeks} accessedsteps < \sum_{allweeks} totalsteps \; x \; 0.8 \qquad (2)$$

Whole course dropout prediction **(CP)**

[4] https://textblob.readthedocs.io/en/dev/

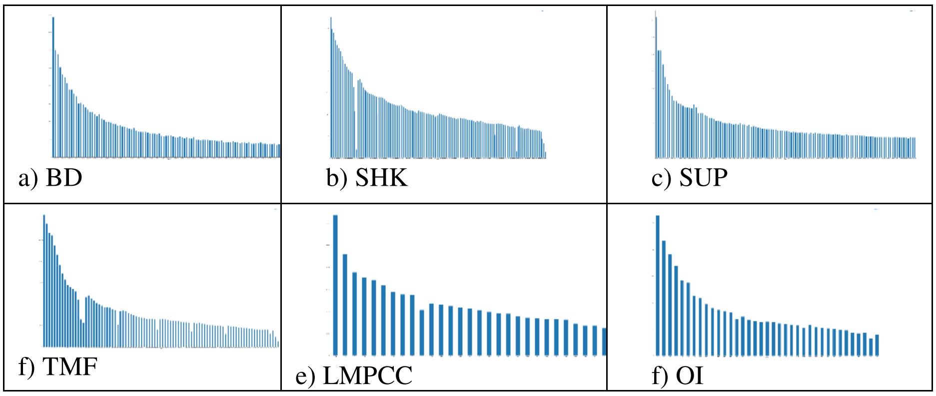

Fig.1 Remaining students over time in different courses (a-f)

Although, about 3-15% of participants complete their courses in MOOC [5], dropout is a gradual process. We are interested in analysing and predicting those weekly dropouts. Figure 1 presents the number of weekly dropouts over time. Clearly, participants are most likely to drop out in the first few weeks. Therefore, identifying those early dropouts is important for prediction. Moreover, using the jumping behaviour feature, we can capture those early dropouts and thus improve the accuracy.

3.3 Proposed Machine Learning Model

We present an overview of the proposed model to predict students' future activities, such as *next week dropout*. The first phase is to clean the datasets, by removing the blank values and missing data. Still, the literature has reported that class imbalance can affect ML algorithms' performance. Due to the massive different completers' ratio to non-completers in our dataset, we set the class weight [5, 28] to the inverse of the frequency of different classes. In terms of best performing learning algorithms, the use of random forest (RF) (e.g., [29, 30, 31, 15]) has appeared in the literature among the most frequently used approaches for the student classification tasks. Additionally, Ensemble Methods, such as boosting, error-correcting have been shown to often perform better than single classifiers, such as SVM, KNN and Logistic Regression [32, 28]. In addition, KNN is an instance-based method, whilst logistic regression is a functional model.

To build our model, we employed several competing ML ensembles methods, as follows: Random Forest (RF) [3], Gradient Boosting Machine (Gradient Boosting), [33] Adaptive Boosting (AdaBoost) [34] and XGBoost [32] to proceed with exploratory analysis. Ensembles refers to those learning algorithms that fit a model via combining several simpler models and converting weak learners into strong ones [26]. In cases of binary classification (like ours), Gradient Boosting uses a single regression tree to fit on the negative gradient of the binomial deviance loss function [24]. XGBoost, a library for Gradient Boosting, contains a scalable tree boosting algorithm, which is widely used for structured or tabular data, to solve complex classification tasks [32]. AdaBoost is another method, performing iterations using a base algorithm. At each interaction, AdaBoost uses higher weights for samples misclassified, so that this algorithm focuses more on difficult cases [34]. Random Forest is a method that uses a number of decision trees constructed via bootstraping resampling and then applying majority voting or averaging to perform the estimation [3].

The current study used a balanced accuracy score (BA) to evaluate the performance of the models; this metric is widely used to calculate accuracy for *imbalanced* datasets, by preventing the majority of negative samples from biasing the result [9]. Moreover, we used the McNemar's [35] test to measure the significance of any improvement in the models after considering student jumping behaviours. Significance levels were set at the 5% level ($P \leq 0.05$).

4 Results and Discussion

This section shows the performance results generated by our seven chosen ML algorithms: Random Forest (RF), Adaboost Classifier (AdaB), XGBoost (XG), GradientBoosting (GBoost), k-nearest neighbour (KNN), Logistic Regression (LR), and extraTrees Classifier. We examine students' learning pattern, accessing time and registration date as mentioned before, for the coming week dropout prediction. Figure 2 shows that participants are more likely to complete the weekly learning activities at the beginning and dropout as time has passed. Around 7500 students have completed the first week in Big Data course. In contrast, only 2223 completed week 5. Therefore, weekly prediction is a reasonable approach to determine at-risk students at an early stage.

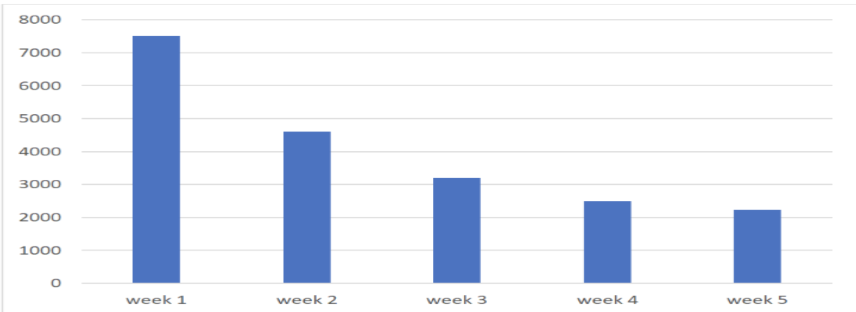

Fig.2 Number of completers students in each week (Big Data course)

4.1 Weekly Prediction

Table 2 shows the performances of models for courses, evaluated by Balanced Accuracy score, a commonly used metric for binary classification of unbalanced dataset. In general, the most robust model is Random Forest (RF), as it outperforms in four courses: 'Supply Chain', 'The Mind is Flat', 'Big Data', and 'Babies in Mind'. Table 2 also shows the performance of several predictive models for both **CP** (whole course dropout, which means if the learner did not access 80% of the topics in the whole course) and **WP** (weekly dropout prediction which means if the learner did not access 80% of the topics in the next week). The input data was extracted from the first week of each course. The two results show that all seven models performed better with weekly predictions using the same number of previous weeks' data and achieved higher accuracy.

The prediction accuracy differences between weekly dropout prediction and the whole course dropout prediction are highlighted in Table 2. The results show clearly how the method of weekly prediction has contributed in terms of accuracy for dropout prediction from early stage (wee k 1). Fig 3 shows the most robust prediction models of weekly drop out and drop out of the whole course.

Table 2. Results (balanced accuracy score (BA)) for prediction models in week 1 for both "weekly dropout prediction" and "dropout from the whole course"

		Testing Balanced Accuracy score (**BA**)						
		AdaB	ExTrees	GBoost	KNN	LR	RF	XG
BIM	CP	50.00%	72.50%	58.13%	59.39%	78.22%	80.32%	54.78%
	WP	**84.48%**	**79.07%**	**67.48%**	**69.50%**	**82.52%**	**83.05%**	**73.59%**
BD	CP	50.00%	78.97%	53.77%	51.47%	86.77%	87.28%	51.44%
	WP	**91.98%**	**90.76%**	**89.31%**	**89.39%**	**92.05%**	**92.03%**	**91.84%**
SHK	CP	50.00%	69.17%	60.47%	61.42%	80.01%	81.02%	59.65%
	WP	**87.15%**	**85.90%**	**81.90%**	**82.79%**	**84.07%**	**87.23%**	**86.97%**
SUP	CP	50.00%	78.61%	67.69%	65.47%	88.53%	86.59%	61.85%
	WP	**93.28%**	**90.26%**	**84.50%**	**81.33%**	**92.09%**	**91.45%**	**91.27%**
TMF	CP	50.00%	74.31%	57.03%	58.77%	85.58%	86.77%	53.98%
	WP	**91.27%**	**90.27%**	**84.77%**	**81.72%**	**90.07%**	**91.43%**	**90.66%**
IMPCC	CP	89.12%	86.34%	78.39%	80.24%	88.12%	86.86%	84.08%
	WP	**90.63%**	**88.98%**	**84.48%**	**85.11%**	**90.12%**	**88.40%**	**90.14%**
OI	CP	90.67%	81.16%	77.50%	71.18%	83.00%	84.07%	78.87%
	WP	**91.41%**	**87.21%**	**77.78%**	**74.75%**	**81.99%**	**85.02%**	**86.53%**

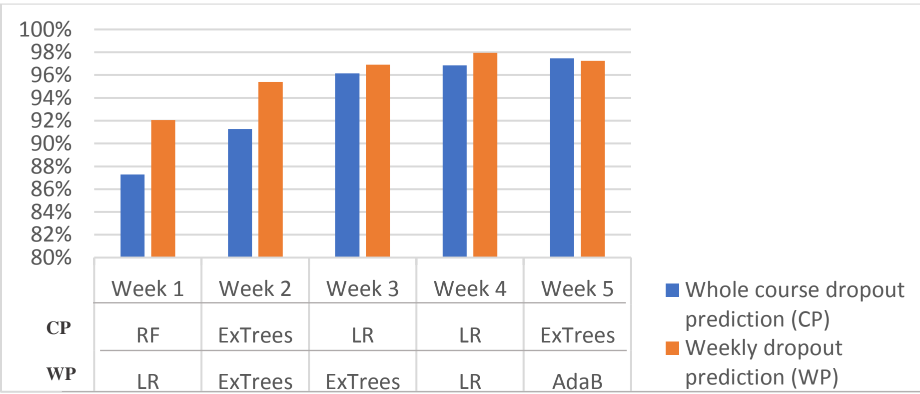

Fig 3: Big data: weekly prediction vs entire course prediction per week with the best performing model

Weekly dropout prediction and the whole course dropout prediction are highlighted in Table 2 and Figure 3. It has been shown how the method of weekly prediction has contributed to the increase in the accuracy of dropout detection from early stage.

4.2 Weekly Prediction with Jumping Activities

We have verified the improved performance of prediction after considering learners' jumping behaviour in four courses. For example, after incorporating the jumping learning pattern as a new feature to the dataset, accuracy rises by nearly 4% - from 86.9% to 91.3% in the XGBoost models in the Shakespeare course. In the Big Data course, the accuracy improves by nearly 3.3%, to 94% for the ExtraTrees Classifier. This weekly dropout prediction improvement is even more generalised in the Open Innovation in Business course, where all seven models implemented are more insightful and the highest accuracy is 94.95%, after considering the jumping learning behaviours. In addition, Table 3 shows that these results were statistically significant between *WP* and *WPWJ* (p value ≤0.05). Based on this analysis, module instructors could implement early interventions, judged on a weekly basis, to improve students' engagement at risk for the upcoming week dropout.

Table 3. Results (BA) of prediction models in week 1 for both weekly dropout prediction (**WP**) and weekly dropout prediction with jumping activities (**WPWJ**)

Testing Balanced Accuracy score (BA)															
Course		**AdaB**	**P. V**	**ExTrees**	**P.V**	**GBoost**	**P. V**	**KNN**	**P.V**	**LR**	**P.V**	**RF**	**P.V**	**XG**	**P.V**
BD	WP	91.98%	P ≤0.05	90.76%	P ≤0.05	89.31%	P ≤0.05	89.39%	P ≤0.05	92.05%	P ≤0.05	92.03%	P ≤0.05	91.84%	P ≤0.05
	WPWJ	**94.73%**		**94.07%**		**92.52%**		**92.01%**		**94.66%**		**94.56%**		**94.68%**	
SH	WP	87.15%	P ≤0.05	85.90%	P ≤0.05	81.90%	P ≤0.05	82.79%	P ≤0.05	84.07%	P ≤0.05	87.23%	P ≤0.05	86.97%	P ≤0.05
	WPWJ	87.15%		**89.74%**		**87.38%**		**86.15%**		**87.58%**		87.16%		**91.32%**	
IM	WP	90.63%	P ≤0.05	88.98%	P ≤0.05	84.48%	P ≤0.05	85.11%	P ≤0.05	90.12%	P ≤0.05	88.40%	P ≤0.05	90.14%	P ≤0.05
	WPWJ	**94.62%**		**93.81%**		**92.12%**		**88.03%**		**94.60%**		**91.29%**		**94.38%**	
OI	WP	91.41%	P ≤0.05	87.21%	P ≤0.05	77.78%	P ≤0.05	74.75%	P ≤0.05	81.99%	P ≤0.05	85.02%	P ≤0.05	86.53%	P ≤0.05
	WPWJ	**94.95%**		**92.76%**		**85.86%**		**79.97%**		**93.27%**		**92.7609**		**90.07%**	

P.V: P-value to show significant difference between WP and WPWJ perfection

Figure 4 shows the Feature importance [26] in Random Forests (our most robust model for this cohort) for the Big Data course. The *Number of Jumping activities* feature is ranked as number one in terms of the importance in predicting students' dropout. However, Figure 4 also shows that the *Number of accesses feature* is the second most important one.

10

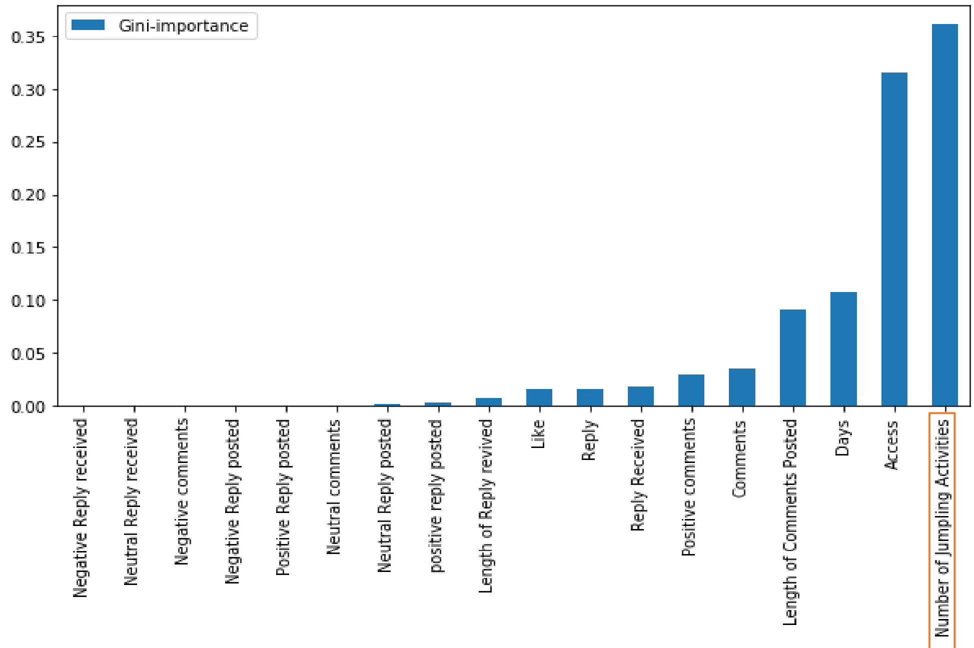

Figure 4: Importance of predictive features

5. Conclusion

This study implements seven predictive models to provide information to enable early interventions for learners at-risk of drop out in the following week from MOOCs. To solve the imbalance dataset problem characteristic for MOOCs (in that successful, completing learners are usually much fewer than non-completers), we set the class weight to the inverse of the frequency of different classes. By reviewing students' learning patterns, particularly *jumping learning behaviours* and previous total course accessing activities, we propose robust machine learning algorithms to build predictive models across seven courses accessed by 251,662 students from 2013 to 2018. Our best model's accuracy (AdaBoost) for the next week dropout learner's detection ranges from 91.41%% to 94.95% in the Open Innovation in Business course, after considering participants' jumping behaviours which could be utilised to personalise and prioritise assistance at-risk learners. Researchers can further add more learners' features (i.e., educational background, age, gender, nationality) to examine further improvements in prediction accuracy in a broad educational context. Additionally, researchers may also deploy the state of art language modelling like Bidirectional Encoder Representations from Transformers (BERT) and XLNet for natural language processing task (Yang et al., 2019) for comment analysis and sentiment analysis in MOOC prediction. Future studies can also explore knowledge representation learning methods based on students' knowledge background.

References

1. Alamri, Ahmed, et al. "Is MOOC Learning Different for Dropouts? A Visually-Driven, Multi-granularity Explanatory ML Approach." International Conference on Intelligent Tutoring Systems. Springer, Cham, 2020.
2. Balakrishnan, G., Co etzee, D.: Predicting student retention in massive op en online courses using hidden markov models. Electrical Engineering and Computer Sciences University of California at Berkeley 53, 57–58 (2013)
3. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
4. Brinton, C.G., Chiang, M., Jain, S., Lam, H., Liu, Z., Wong, F.M.F.: Learning about social learning in moocs: From statistical analysis to generative model. IEEE transactions on Learning Technologies 7(4), 346–359 (2014)
5. Coates, A., Carpenter, B., Case, C., Satheesh, S., Suresh, B., Wang, T., Wu, D.J., Ng, A.Y.: Text detection and character recognition in scene images with unsupervised feature learning. In: 2011 International Conference on Document Analysis and Recognition. pp. 440–445. IEEE (2011).
6. Drummond, C., Holte, R.C., et al.: C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In: Workshop on learning from imbalanced datasets II. vol. 11, pp. 1–8. Citeseer (2003)
7. Gardner, J., Bro oks, C.: Student success prediction in moo cs. User Modelling and User-Adapted Interaction 28(2), 127–203 (2018)
8. Fox, C.: Futurelearn has 3 million learners (Mar 2016), https://www.future-learn.com/info/press- releases/futurelearn- has- 3- million- learners
9. Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. In: 2010 20th international conference on Pattern recognition (ICPR), pp 3121–3124. IEEE
10. Hong, B., Wei, Z., Yang, Y.: Discovering learning behavior patterns to predict dropout in mooc. In: 2017 12th International Conference on Computer Science and Education (ICSE). pp. 700–704. IEEE (2017)
11. Jeon, B., Park, N.: Dropout prediction over weeks in moocs by learning representations of clicks and videos. arXiv preprint arXiv:2002.01955 (2020)
12. Jordan, K.: Massive open online course completion rates revisited: Assessment, length and attrition. International Review of Research in Open and Distributed Learning 16(3), 341–358 (2015)
13. Khalil, H., Ebner, M.: Moocs completion rates and possible methods to improve retention-a literature review. In: EdMedia+ innovate learning. pp. 1305–1313. Association for the Advancement of Computing in Education (AACE) (2014)
14. Kloft, M., Stiehler, F., Zheng, Z., Pinkwart, N.: 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. pp. 60–65 (2014)
15. Liang, J., Li, C., Zheng, L.: Machine learning application in moocs: Dropout prediction. In: 2016 11th International Conference on Computer Science & Education (ICCSE). pp. 52–57. IEEE (2016)
16. Lu, X., Wang, S., Huang, J., Chen, W., Yan, Z.: What decides the dropout in moocs? In: International Conference on Database Systems for Advanced Applications. pp. 316–327. Springer (2017)

17. Pereira, F.D., Oliveira, E.H., Oliveira, D.B., Cristea, A.I., Carvalho, L.S., Fonseca, S.C., Toda, A., Isotani, S.: Using learning analytics in the amazonas: understanding students' behaviour in introductory programming. British journal of educational technology 51(4), 955–972 (2020)
18. Prati, R.C., Batista, G.E., Monard, M.C.: Learning with class skews and small disjuncts. In: Brazilian Symposium on Artificial Intelligence. pp. 296–306. Springer (2004)
19. Ren, Z., Rangwala, H., Johri, A.: Predicting performance on mooc assessments using multi-regression models. arXiv preprint arXiv:1605.02269 (2016)
20. Robinson, C., Yeomans, M., Reich, J., Hulleman, C., Gehlbach, H.: Forecasting student achievement in moocs with natural language processing. In: Proceedings of the sixth international conference on learning analytics & knowledge. pp. 383– 387 (2016)
21. Rose, C., Siemens, G.: Shared task on prediction of dropout over time in massively open online courses. In: Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs. pp. 39–41 (2014)
22. Sunar, A.S., White, S., Abdullah, N.A., Davis, H.C.: How learners' interactions sustain engagement: a mooc case study. IEEE Transactions on Learning Technologies 10(4), 475487 (2016)
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
24. Wen, M., Yang, D., Ros´e, C.: Linguistic reflections of student engagement in massive open online courses. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 8 (2014)
25. Xing, W., Chen, X., Stein, J., Marcinkowski, M.: Temporal predication of dropouts in moocs: Reaching the low hanging fruit through stacking generalization. Computers in human behavior 58, 119–129 (2016)
26. Yang, D., Sinha, T., Adamson, D., Ros´e, C.P.: Turn on, tune in, drop out: Anticipating student drop outs in massive op en online courses. In: Pro ceedings of the 2013 NIPS Data-driven education workshop. vol. 11, p. 14 (2013)
27. Yang, Z., Dai, Z., Yang, Y., Carb onell, J., Salakhutdinov, R., Le, Q.V.: Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237 (2019)
28. Dietterich, Thomas G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45014-9_1
29. Sharkey, M., Sanders, R.: A process for predicting MOOC attrition. In: Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs, pp. 50–54 (2014)
30. Nagrecha, S., Dillon, J.Z., Chawla, N.V.: MOOC dropout prediction: lessons learned from making pipelines interpretable. In: International World Wide Web Conferences Steering Committee Proceedings of the 26th International Conference on World Wide Web Companion, pp. 351–359 (2017)
31. Bote-Lorenzo, M.L., Gómez-Sánchez, E.: Predicting the decrease of engagement indicators in a MOOC. In: Proceedings of the Seventh International Learning Analytics and Knowledge Conference on LAK 2017. pp. 143–147. ACM Press, New York (2017)
32. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, ACM. (2016)
33. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232 (2001)
34. Hastie, T., Rosset, S., Zhu, J., Zou, H.: Multi-class adaboost. Statistics and its. Interface 2, 349–360 (2009)

35. B. Everitt, The Analysis of Contingency Tables, Chapman and Hall, London, 1977.
36. Pereira, F.D., Oliveira, E., Cristea, A., Fernandes, D., Silva, L., Aguiar, G., Alamri, A. and Alshehri, M., 2019, June. Early dropout prediction for programming courses supported by online judges. In International Conference on Artificial Intelligence in Education (pp. 67-72). Springer, Cham.