Article

A Case Study on English as a Second Language Speakers for Sustainable MOOC Study

Ismail Duru [1], **Ayse Saliha Sunar [2,]***, **Su White [3]**, **Banu Diri [1]** and **Gulustan Dogan [4]**

1   Department of Computer Engineering, Yildiz Technical University, Istanbul 34220, Turkey;
     iduru@yildiz.edu.tr (I.D.); diri@yildiz.edu.tr (B.D.)
2   Department of Computer Engineering, Bitlis Eren University, Bitlis 13000, Turkey
3   Department of Electronic and Computer Science, University of Southampton, Southampton SO17 1BJ, UK;
     saw@ecs.soton.ac.uk
4   Department of Computer and Instructional Technologies, Yildiz Technical University, Istanbul 34220, Turkey;
     gulustan@yildiz.edu.tr
*   Correspondence: assunar@beu.edu.tr

Abstract: Massive Open Online Courses (MOOCs) have a great potential for sustainable education. Millions of learners annually enrol on MOOCs designed to meet the needs of an increasingly diverse and international student population. Participants' backgrounds vary by factors including age, education, location, and first language. MOOC authors address consequent needs by ensuring courses are well-organised. Learning is structured into discrete steps, prioritising clear communication; video components incorporate subtitles. Variability in participants' language abilities inevitably create barriers to learning, a problem most extreme for those studying in a language which is not their first. This paper investigates how to identify ESL participants and how best to predict factors associated with their course completion. This study proposes a novel method for automatically categorising (English as Primary and Official Language; English as Official but not Primary Language; and English as a second Language groups) 25,598 participants studying FutureLearn "Understanding Language: Learning and Teaching" MOOC using natural language processing. We compared algorithms' performance when extracting discernible features in participants' engagement. Engagement in discussions at the end of the first week is one of the strongest predictive features, while overall, learner behaviours in the first two weeks were identified as the most strongly predictive feature.

Keywords: English as a second language; MOOCs; natural language processing; prediction model; sustainability

1. Introduction

MOOCs are increasingly recognised as providing high-value learning resources enabling an accessible route to sustaining the expansion of both formal and informal education. Formally, blended MOOCs provide a means whereby academics can incorporate externally produced resources into their face-to-face teaching. This is being used as a means to rapidly expand and grow educational capacity—for example in women's universities in Saudi Arabia [1]. Informally, individual learners typically use MOOCs to access educational independently and to update professional skills and gain access to education for little or no cost. However the use of MOOCs in developing countries is not as straightforward as some might assume with language barriers having been identified as one of the important problem areas [2]. The focus of this paper working on methods to automatically identify different types of learners according to their language background, and then looking at means to predict likely learning pathways is particularly relevant to these contexts.

In 2018, according to ClassCentral (https://www.class-central.com/report/mooc-stats-2018), 100 million people around the world studied 11.4K Massive Open Online Courses (MOOCs) delivered by over 900 institutions. Of these, 20 million individuals enrolled on a course for the first time, slightly less than the 23 million first-time enrolled learners the previous year. The top five MOOC providers did not change in 2018. The majority were English language based—US providers Coursera (37M); edX (18M); Udacity (10M) plus the UK provider FutureLearn (8.7M); China's XuetangX accounted for 14M learners.

Even though most MOOCs are in English and most providers are US based, the demographics of participants are diverse. According to one MIT's report, 71% of their participants are international (https://news.mit.edu/2017/mooc-study-offers-insights-into-online-learner-engagement-behavior-0112).

MOOCs are considered as a mean of democratising the education as MOOCs openly and freely provides the materials to anyone having the Internet connection and a proper device. However, Dillahunt et al. [3] pointed out numbers of aspects proving the reality is far different. The authors also expressed that the use of English as the primary language of instruction in MOOCs needs to be discussed. The large proportion of international participants raises the issue on possible difficulties with communicating in English for those to whom English is their second language. One solution to overcome such barriers has been developed by institutions such as EMMA (https://platform.europeanmoocs.eu/) to deliver MOOCs in multiple European languages. Mamgain et al. [4] highlight how some MOOC providers (e.g., Coursera) provide learners with English transcript option and subtitles in different languages. In one study, undertaken by Eriksson et al. [5], MOOC learners stated that such subtitles in English were helpful.

One important indicator suggesting that a learner is likely to succeed in a course, is their engagement in discussion forums through conversations, because it requires some level of English language fluency. Cho and Byun [6] reported that there is very little evidence on how English as a second Language (ESL) participants study in MOOCs. As Kizilcec et al. [7] explain, people from Least Developed Countries where English is typically not the primary language, may feel discouraged because of their fear of being seen as less capable through their poor language skills. This factor may cause them to contribute less to conversations, resulting in reduced engagement with the course and a greater chance that students may drop the course.

A potentially valuable contribution to this area may lie in analysing and modelling the behaviour of ESL participants. This data and the insights gained from the analysis can then be used as a basis for a prediction model. A recommender system, and other design enhancements based on such a model could then be used to help ESL participants for their MOOC study. We are interested in investigating in greater detail the behaviours of second language learners in MOOCs. We also aim to develop a prototype tool to enhance the learning experience of this demographic group.

Uchidiuno et al. [8] interview international students to understand their motivation in the context of accessibility of MOOCs. The authors suggest that translating the content could be a solution but it may not be suitable for everyone. However, tailored tools, based on the needs and motivation of English as a second language speaker (ESL) participants could be more effective.

In another example, Colas et al. [9] show that discussion forums in additional languages result in an improvement in engagement being observed. However, their method, recruiting seven mentoring teams for monitoring the discussions, is not extensible. Additionally, their case study focused on a narrower perspective of MOOCs [9]. We investigated user behaviours from wider perspectives beyond the MOOC.

It is very important for MOOC providers to understand the needs of their participants to sustain the learning activities on the platform. This work fills in this important gap in MOOC research that can be useful for MOOC providers to identify different groups of English language learners according to their primary languages. Additionally, it provides a detailed comparative analysis on behaviours and performances of learners in different English language groups which has not been addressed by

any study yet to the best of our knowledge. To clarify the novelty of this paper, the contributions of the study are summarised as follows:

1. A novel method with regex patterns is proposed for identifying if participants give any information regarding their nationality, hometown and first language. According to the results, we grouped the participants whether or not they are English as Second Language (ESL).
2. Unlike existing research, the behaviour in course engagement of the participants categorised by their first language is analysed. It is observed that participants are from English as an Official and Primary Language (EPL) group are more actively engaged in the course and are more likely to complete the course.
3. Which behaviours are more predictive for the participants in different language groups are identified. For example, the total steps of Week 1 completed by EPL participants is the most predictive feature while the number of steps is the most predictive for ESL participants.
4. The differences between the algorithms for weekly prediction and for the prediction at the end of the course are compared. The Random Forest Method performed better.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive state-of-the-art literature on the English as a second language participants in MOOCs. Section 3 explains the methodology of the study. Our novel method for automatic identification of English as a second language participants is proposed in Section 4. Course engagement analysis have been carried out in Section 5, including from three aspects: behaviours in course steps in Section 5.1, behaviours in discussion forums in Section 5.2, and behaviours in follow relationships in Section 5.3. Then, the implementations and results of weekly and overall predictions of course completion is explained in Section 6. Finally, the results are discussed and concluded in Section 7 and future work is presented in Section 8.

2. Related Work

2.1. Learning Analytics in MOOCs

A large amount of work with the learning analytics community gather evidence in a way which we think may be relevant to our research contribution. Learning Analytics introduces a systematic approach to understanding student behaviour and performance in MOOCs. Using learning analytics enables rapid development of the measurement, analysis and prediction of student behaviour and performance [10].

Ramesh et al. [11] point out that understanding student engagement as a course progresses helps characterise learning patterns and thus can help minimise dropout rates by prompting focused instructor intervention. A study conducted by Papadakis et al. [12] shows that the presentation of the content and affordances of the platform is also an important factor which may cause reduced engagement. For example, improving mobile application of the course and gamificiation elements to facilitate instructor intervention could have an impact on use for English as a second language speakers. Klusener et al. [13] use learner profiles derived from forum activity to analyse learning behaviour and establish methods which can help identify necessary interventions. Brinton et al. [14] investigated factors that are associated with decreased forum participation and ranked the threads accordingly. They suggested that the identified factors and thread ranking mechanism might be used to make individualised recommendations in MOOC.

In another study, DeBoer and Breslow [15] indicate that each click a participant makes in a MOOC is a part of the online behaviours that help to predict their learning processes and the learner's attitudes towards the course. Other studies ([13,16–19]) focus on learners' performance prediction with MOOCs.

2.2. Engagement in MOOCs from Language Perspectives

In this paper, we particularly focus on the performance of second language English speakers.

Researchers have been investigating the use of learning analytics to gain insight into learners' engagement, detecting needs, and suggesting possible design changes. However, there is limited research investigating behaviours of second language English speakers within MOOCs.

The available studies can be divided into four main themes:

1. Participants' engagement on a MOOC which is delivered in a language other than English [20–23].
2. English Language Learners' (ELL) engagement on a MOOC which is delivered in English ([24–26]).
3. English as a second Language Learners' (ESL) engagement on a language MOOC which is authored for learning English language [27–30].
4. English as a second Language Learners' (ESL) engagement on a MOOC which is delivered in English ([6,31–34]).

One study of Uchidiuno et al. [25] investigates the engagement of English Language Learners participating in MOOCs delivered in English. They find that, in these cases, even though English is not the first language of such language learners, they are professionally interested in the language and the motivation. For this reason, they argue there is a need for the needs of English language learners to be specifically investigated.

A Further study by Uchidiuno et al. [26] suggest that English language learners when compared with those for whom English is a first language show increased interaction with text content and less interaction with video and content without visual support.

Cho and Byun [6] study English as a Second Language Learners' engagement on a MOOC delivered in English. They investigate the experience of 24 Korean college students. They identify language as a potential barrier for active participation amongst ESL learners. Additionally, culturally unfamiliar teaching and learning practices present some difficulties for ESL learners.

Rimbaud et al. [35] draw attention to lack of adaptive MOOC to support second language English speakers in MOOCs. They suggest that "Content and Language Integrated Learning (CLIL)" could be a solution for the needs of ESL learners. De Waard and Demeulenaere [36] integrate MOOCs and CLIL method in a blended learning environment to increase language, and social and online learning skills for 5th grade K-12 students. These provide some evidence that an adaptive support for ESL learners could be beneficial.

Reilly et al. [37] tested an automated essay grading systems and observed that ESL learners are disadvantaged because the scores given by an automated grading system as significantly lower than those of human instructors. The authors recommend that MOOCs should address this disadvantage and take measures for their multicultural and linguistically diverse audience.

Overall researchers are concerned about ESL MOOC participants' needs and motivations and how they engage with content. However, further deeper investigation is required to have a greater understanding of ESL participants' needs.

3. Research Aim and Methodology

The ultimate aim of this research is to identify a reliable method by which we can predict the course completion performance of different language-based groups. By using the results of behavioural analysis of the groups, the findings could be valuable for identifying study strategies for the learners to sustain their engagement in MOOCs.

The research has been designed as a case study which performs a series of analyses by using the data generated from an Understanding Language: Learning and Teaching MOOC on FutureLearn. Figure 1 shows the steps in the operation of the research. After dividing the learners into the groups based on whether or not their primary language is English, the learning analytics techniques are applied to investigate the behaviours of learners in each group. Then, a prediction model is developed based on the findings from the analyses. As a final step, some study strategies will be identified for the MOOC participants, especially for those who speak English as a second language. In this paper, the steps highlighted with red (Step 1, Step 2.1 and Step 3) in the Figure 1 are explained.
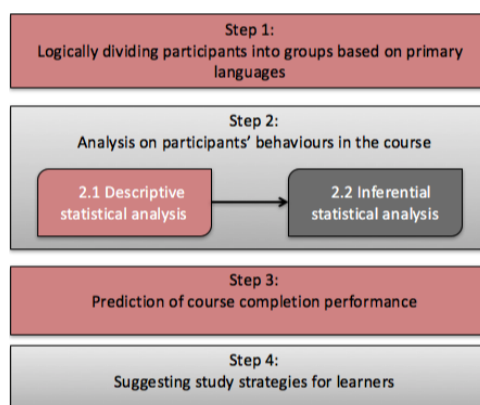
Figure 1. Four steps to identify effective study strategies.

We investigated the following research questions:

- Is it possible to automatically identify the English as a second language speakers from comments in discussion forums?
- Is there any difference between the behaviours of English as a second language participants and the other participants in completing the course, contributing to the discussions, and interacting with each other?
- If there is, is it possible to use these differences to build a predictive model and predict the participants' completion of course?

Datasets

The FutureLearn dataset from a four-week course named "Understanding Language: Learning and Teaching", that ran between 4 April 2016 and 2 May 2016 was used. We analysed the total dataset up to and including the final day 14 May 2016. We chose this course since the MOOC is about learning language and naturally conversations are built around languages. Also, many international English language teachers enrolled in the course. This range of diverse participants produce a rich source of data where we might detect who is an English as a second language speaker.

The dataset includes the following data files:

- **Enrollments:** Demographic information for each participant including enrol/unenrol/purchase certificate date.
- **Step Activity:** Participant data for each step (learning unit) page opened and completed.
- **Comments:** All posts in the discussion forums (comment author, content, when, to whom was directed).
- **Followings:** All following relations (who followed whom and when).

4. Categorising MOOC Participants Based on First Languages

In our previous study [38], we detected that some participants declared that their first language was different from the official language(s) of the country they live in.

A participant's primary language may be identified via demographic data gathered from registration surveys. However, participants frequently do not answer the questionnaires.

Various approaches have been suggested to overcome this, for example Uchidiuno et al. [39] detect participants' primary languages via their language preferences on their browsers. However, there is always a possibility that a learner may have chosen English as a browser language preference for practicing purposes.

In our research, in order to group the participants according to their language, we take into consideration (i) the location information they gave during enrollment and (ii) their statements in the discussion forums using a computational rather than manual approach.

Of 25,597 enrolled people, 3305 participants (12.9%) gave location information during the enrollment process. We assumed that the first language of a person is the same with the language of the country where they live unless they stated otherwise in their conversation on the discussion forums.

Consequently, we grouped these 3305 participants in three language-based groups:

1. *English as an Official and Primary Language (EPL)*: Participants who stated in the pre-course survey or in discussions that they are from a country where English is the official and primary language e.g., the United Kingdom. Also participants who stated in their posts that their first language is English.
2. *English as an Official but not Primary Language (EOL)*: Participants who stated in the pre-course survey or in discussions that they are from a country where English is one of the official languages but not the primary language e.g., India.
3. *English as a Second Language (ESL)*: Participants who stated in the pre-course survey or in discussions that they are from a country where English is neither a primary language nor an official language e.g., Turkey, also participants who stated in their posts that their first language is not English.

Automatic Detection of ESL Participants with Natural Language Processing (NLP)

Our previous approach, manually going through all discussion threads was very time consuming and not scalable [38]. The new approach exploits regular expression patterns and similarity metrics to identify the information about country, city, nation, and language in the discussion. We then process the information to categorise whether or not the first language of a participant is English.

Detecting patterns: As a first step, we detect which words and phrases a participant uses to talk about their first language. We detected several different types of sentences that include location, language, or nationality information. We generated 22 regex patterns to detect these types of sentences. Here are some sample examples from the discussion forums used to shape the patterns:

- I **was born** in **Turkey**.
- I am a **native speaker** of **Spanish**.
- **My mother's tongue** is **Albanian**.
- I live in the UK but I **am** actually **French**.
- I **am** a typical **Japanese**.
- I only **speak English**.
- My **first tongue** is **English**.
- I **grew up** in **Australia**.

Formalising patterns: We then formalised the patterns as regular expressions (regex codes). One regex code (and one pattern) could match more than one sample sentence. For example, the pattern

$$\text{"I } was \mid were \text{ brought } up \mid born \mid grow \text{ } up \mid grew\text{"}$$

$$\text{" } \mid grown \mid grown \text{ } up \text{ } country \mid city\text{"}$$

is formalised with regex codes. The algorithm matches the sentences "I was born in Turkey" and "I grew up in Paris" with this pattern.

Extracting the real language information from discussion posts: When the algorithm processes sentences in discussions, it matches the sentences with the patterns and extracts the information related to country, city, nationality, and language. Using this evidence, the algorithm defines the language and/or country information of participants.

Comparing and refining the language information: Once the algorithm identifies the country/language information of participants, it compares this information with the location information of participants which is already in the database. Language-based group information

extracted from the post might not correspond with the location information of the learner in the database. In this case, the language-based group information is updated according to the discussion post.

Figure 2 shows the system model for identifying which language group a participant belongs to by using the survey information and the discussion forums. When a participant posts a comment including information about their primary language, nationality, or country they live in, it is possible for us to identify this information from comment text with a regex method.
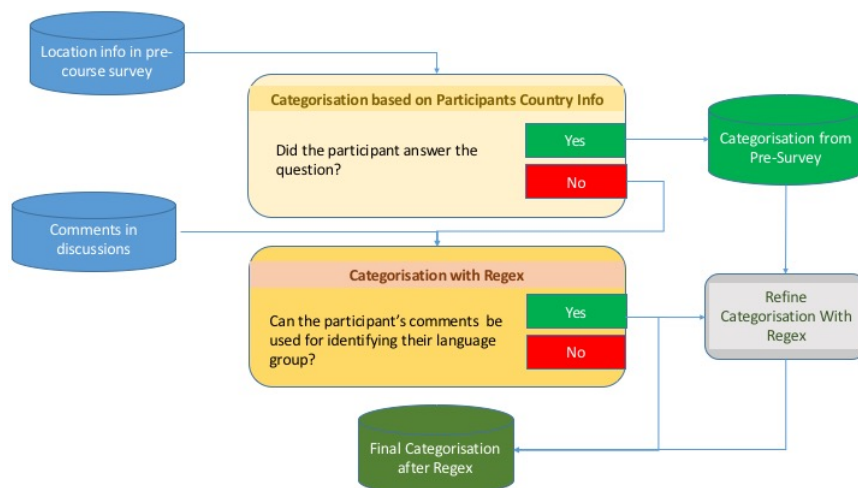
Figure 2. The proposed model for language group categorisation using regex.

Table 1 shows that 3305 participants filled in the pre-course survey indicating where they were participating from. Additionally, 674 (20.40%) of them chose to share their personal information such as where they are from, what their nationality is, which language their first language is, and what their nationality is in the discussions forums. Analysing their statements in discussions and their responses in the pre-course survey, our algorithm identifies that:

- 643 (19.46%) participants are first language English speakers;
- 434 (13.13%) participants live in a country where English is the primary language but their first languages are not English;
- English is the second language of 2228 (67.41%) participants.

Table 1. Stages in detecting and refining language-based group of participants.

I. Participants who responded the question about where they attend the course from			
Total	**EPL**	**EOL**	**ESL**
3305	602 (18.21%)	459 (13.89%)	2244 (67.90%)
II. Participants who gave additional information regarding to their language/country in discussion, apart from the pre-course survey			
Total	**EPL**	**EOL**	**ESL**
674	148 (21.96%)	68 (10.09%)	458 (67.95%)
III. Refine the assumed language group with the extracted information from discussions			
Total	**EPL**	**EOL**	**ESL**
3305	643 (19.46%)	434 (13.13%)	2228 (67.41%)

5. Course Engagement Analysis

This section presents statistical analysis of how learners in different language-based groups interact with the course e.g., how they use discussion forums, how often they complete the course steps. We mainly focus on behaviours in completing course steps, in attending discussion forums, and in following others.

The analysis step of the research has been conducted to answer the second research question: Is there any difference between the behaviours of English as a second language participants and the other participants: In completing the course, contributing to the discussions, and interacting with each other? The data generated by FutureLearn allows us to track participants' engagement with

- the course steps i.e., when a learner opened the page of the step and marked as completed;
- the discussion forum i.e., what a learner posted to discussion threads;
- the followings feature i.e., whom a learner started following another learner.

Therefore, we have divided the section into three subsections to analyse the above mentioned behaviours respectively.

5.1. Behaviours in Course Steps

The engagement level of different MOOC learners varies. However, a trend of decreasing overall engagement by course participants over successive weeks is widely observed in almost every MOOC. Clow [40] described this behavioural pattern as *Funnel of Participation*. Analysing the pattern of participation in the course confirmed this decreasing trend. Figure 3 shows a pattern of decreasing retention for each language-based group. The highest percentage of participation in course steps and course completion is observed in the EPL learners.
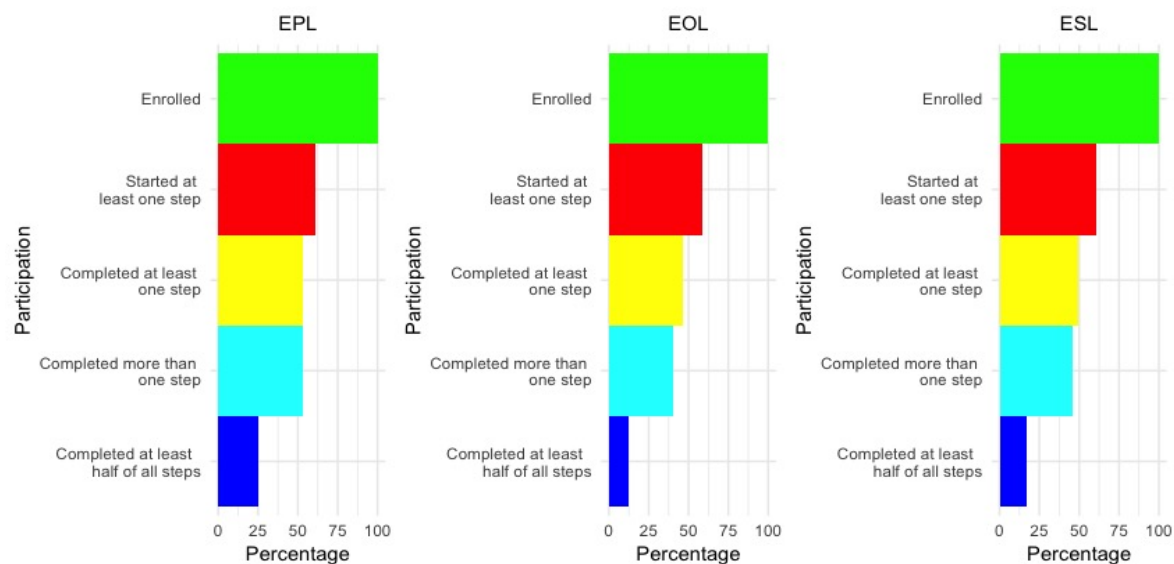

Figure 3. Analysing and comparing participation within each language group.

The violin plot in Figure 4 shows the overall completion ratio of steps at the end of the course by the three defined language-based groups. The ratio is calculated from the percentages of the completed steps among the all steps in the course, not only the steps a learner started. Violin plots show the probability density, median value, and interquartile range and are useful to present comparisons of sample distributions across different categories.
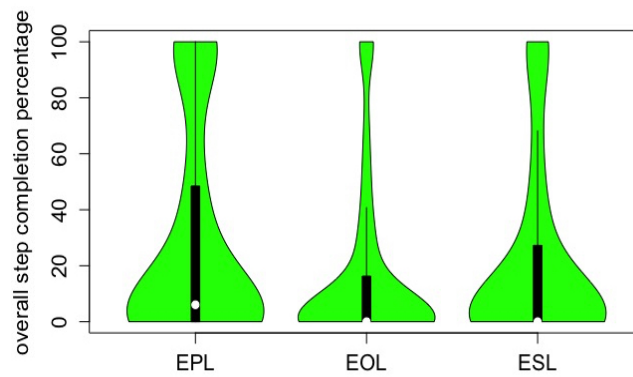
Figure 4. Overall step completion percentage by different language-based groups.

Figure 4, shows that the overall course completion ratio of EPL participants is higher than any other language-based groups.

5.2. Behaviours in Discussion Forum

A learner can make a contribution to discussions by posting an original comment, replying to a comment, or liking a comment/reply. Social attendance via these actions is even smaller than the overall course participation.

In one of our previous studies [41], we conducted a narrower investigation on forum behaviours of language oriented MOOC learner groups. We found that there is a significant difference between each language group in terms of number of weekly comments and replies posted and the popularity of the comments posted across the different language groups.

The pie charts in Figure 5 show the ratio of learners who are active or passive participants in discussion forums showing little difference in the volume of contributions to the discussions across language groups. Very similar proportions of each group wrote at least one original comment or a reply to the discussions.

Figure 6 indicates the number of total comments, including original comments and replies per participant, that are posted by learners in different groups. According to Figure 6, EPL participants posted a larger number of comments to the discussions. There are some outliers who posted more than 400 comments. However, the majority of the participants in each language-based group did not post any comments. Across the three language-based groups, EOL participants posted the least number of comments.
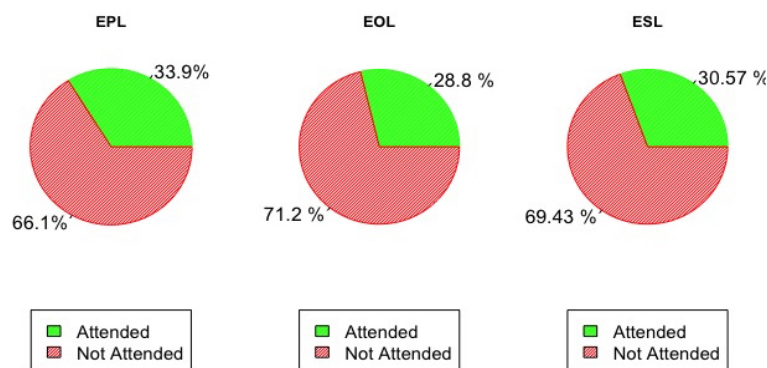

Figure 5. Overall attendance in discussion forums by different language-based groups.
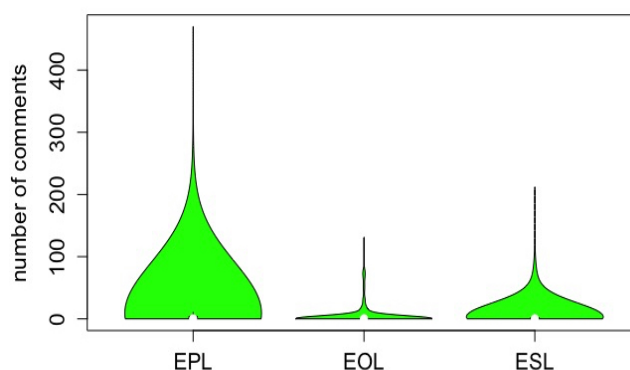
Figure 6. Number of comments posted to discussions by each language-based groups.

In order to accomplish a deeper analysis, we also investigated how many replies and likes attracted a comment. The scatter plot in Figure 7 maps the total number of replies that are attracted against the total number of original comments posted by each learner. The number of original comments and replies is capped at 70 for a clearer illustration in the Figure 7 (Since the data is very large, including outliers for representation makes the graph very small and unclear. For a clearer representation, we have excluded the examples with comments bigger than 70). The trends in the graphs are very similar. No matter how many original comments are posted by a learner, they usually get fewer than 10 replies.
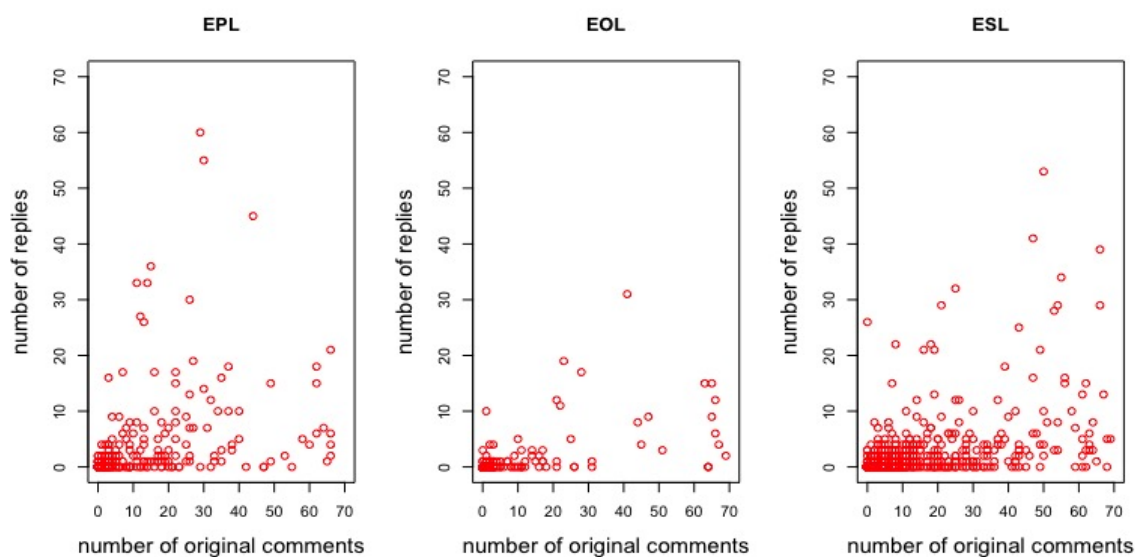

Figure 7. Number of replies received and number of original comments posted by learners.

The scatter plot in Figure 8 shows the total number of likes that are attracted by comments posted by each learner. The maximum number of comments and likes has again been capped to 70 for a clearer illustration.

Figure 9 plots the number of replies received to any original comment posted by learners in each language-based group.

It is observed that the comments posted by EPL learners got more replies than others. However, the majority of all comments typically receive no replies from any group.

Despite the tendency in receiving replies (Figure 7), there is a linear correlation between the number of comments posted and the number of likes received in each language-based group. Figure 10 plots the number of likes attracted by the comments posted by learners in each language-based group. The same behaviour pattern is observed in Figure 6. Even though the majority of comments posted

by learners in each group did not receive any likes, there are learners in EPL who did receive large number of likes. Some learners in ESL also received some large number of likes but it remained less than EPL. No pattern was detected in the way in which EOL participants received likes.
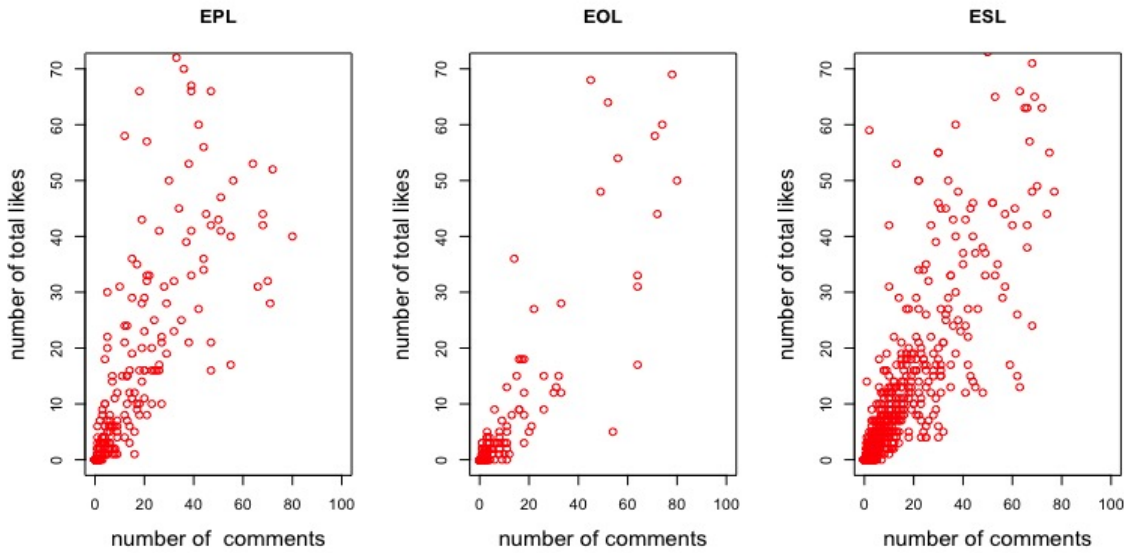

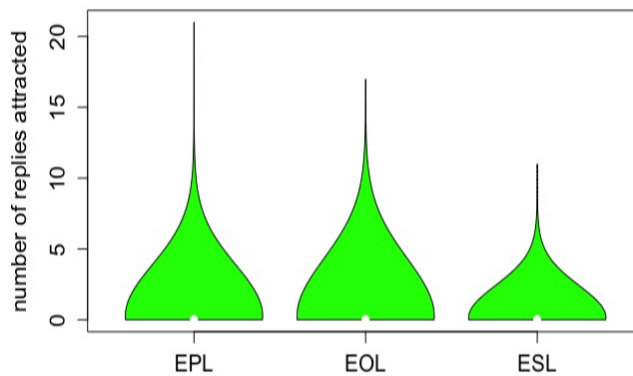**Figure 8.** Number of likes received and number of comments posted by learners.

Figure 9. Number of replies that are attracted to the comments.
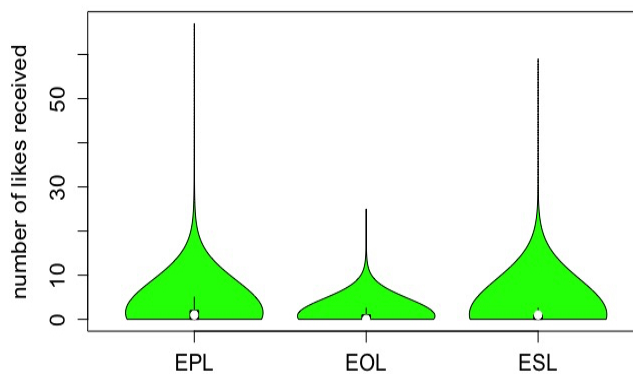

Figure 10. Number of likes that are received.

We also investigated character length of the comments posted by learners in the different language-based groups. Figures 11 and 12 illustrate the average length (indicates the average number of characters in an initial comment or reply) of the initial comments and the replies posted by learners.

To clearly show the difference between language groups which vary considerably in actual size, we randomly selected equal number of learners from each language-based group (Figures 11 and 12).

The results indicate that ESL participants sometimes posted initial comments longer than 800 characters, however, ESL and EOL participants mostly post less than 200 characters while EPL learners mostly post more than 200 characters.

In the case of replies (Figure 12), the difference between language-based groups is more distinct. EOL participants' replying comments are rarely longer than 100 characters. ESL participants replying comments are rarely longer than 200 characters. Contrarily, EPL participants' replies typically longer than 100 characters. Also, EPL learners more frequently give replies to comments than any other groups.
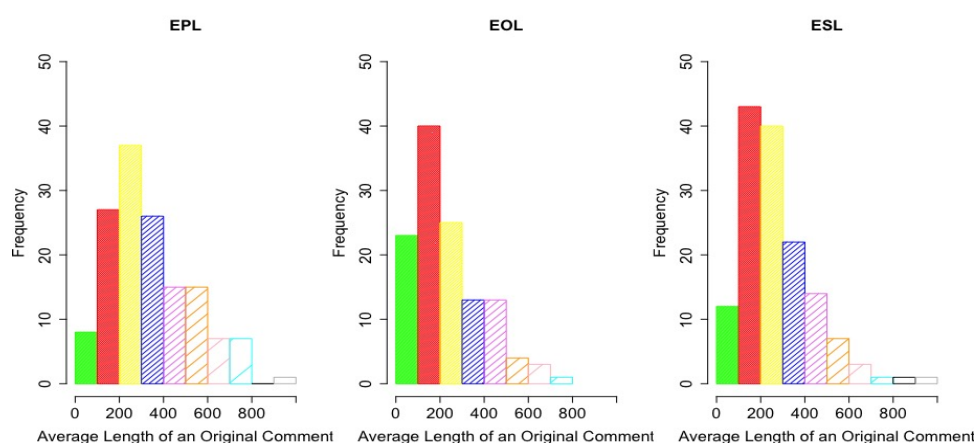

Figure 11. Lengths of original comments posted by participants (for equal number of learners randomly selected from each language group).
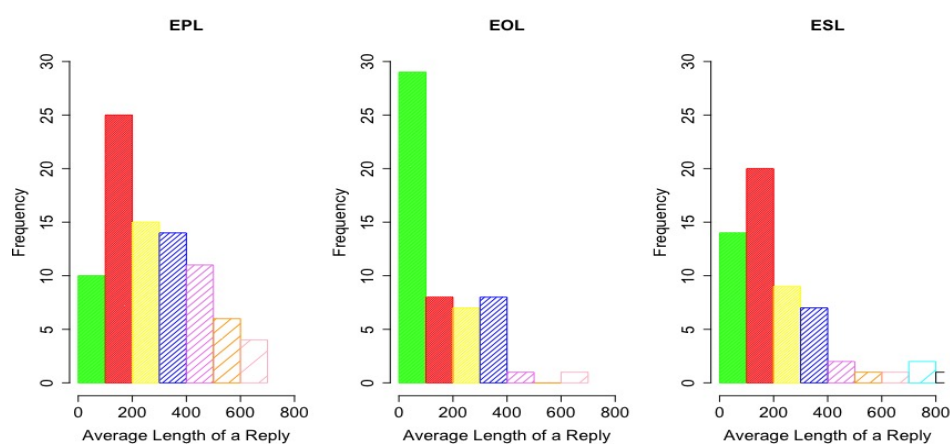

Figure 12. Lengths of replies posted by participants (for equal number of learners randomly selected from each group).

5.3. Following Behaviours

Previous research conducted by Sunar [42] found that following behaviours positively correlated to step completion in a course, particularly when the learners participated in conversations by posting comments. We have found a similar behaviour pattern in this analysis, however, there are differences in the details between the different language-based groups.

Figure 13 shows how many fellow learners are followed by each learner in the different language-based groups from the same course. None of the EPL learners follows more than 45 people.
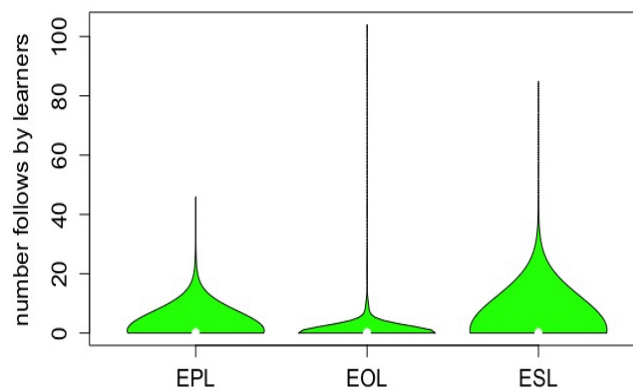
Figure 13. Number of people that participants in each group follow.

According to Figure 13, individuals in EPL and ESL groups generally prefer to follow a greater number of learners. The distribution of EPL learners' behaviour is much more consistent than others. There is a great deal of variability in the 'following' behaviour of ESL group. EOL participants have a fairly coherent subset of behaviours (a small number of EOL learners followed). Both EOL and ESL have a large number of outliers.

Figure 14 indicates how many people were followed by each of the learners in the different language-based groups from the same course. A noticeable difference across groups was observed. Participants from the EPL group were followed by the greatest number of people. EOL learners are almost never followed by fellow learners. ESL learners were typically followed by fewer than 20 learners. This might be because ESL and EOL participants also follow fewer learners than EPL participants.
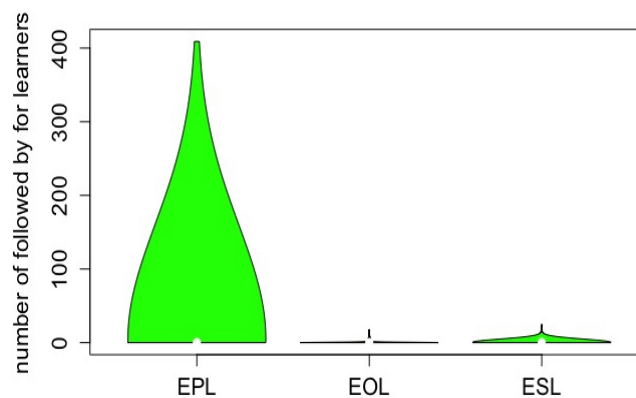

Figure 14. Number of people in each group who are followed by other participants.

The pie charts in Figure 15 show the proportions of followed participants according to their language. For each group, ESL learners had the greatest number of followers. Most learners are followed by EPL and ESL learners, very few from EOL learners. However, this result does not show that ESL learners are likely to use the follow feature because the number of ESL learners are far more than the others. For each group, the number of EOL learners who follow someone is the smallest proportion.
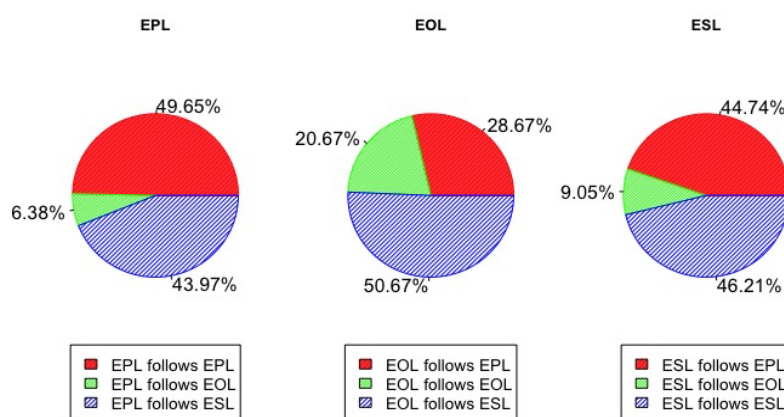
Figure 15. Proportions of followed participants according to their language-based groups.

6. Course Performance Prediction

It can be concluded from the results presented in Section 5 that there are some differences in the performance of language-based groups and the important features that affect their performance.

In this section, we analyse participants' performances and investigate which features might be important in order to predict their course outcomes.

6.1. Feature Extraction

We used data mining techniques (with the R programming language) to extract features from the data which we generated. The extracted features are categorised in three areas: (1) Step Activity, (2) Comments, (3) Followings. Table 2 lists the extracted features in each category.

Table 2. Extracted features with datasets.

1-Step Activity	2-Comments	3-Followings
1. number of steps opened	1. number of comments posted	1. number of fellow participants following a learner during the course
2. number of steps opened **but** not completed	2. number of comments initiated	2. number of fellow participants that a learner follows during the course
3. total completed Week 1 steps	3. number of replies	
4. total completed Week 2 steps	4. average length of all comments	
5. total completed Week 3 steps	5. average length of initiated comments	
6. total completed Week 4 steps	6. average length of replies	
	7. average numbers of likes that a comment received	

This data was then investigated in terms of their relationship with students course performances.

6.2. Balancing Data

We defined the level of course completion across the three categories as follows:

- **Dropouts:** Learners who officially dropped the course or those who completed none of the steps
- **Slow paced learners:** Learners who completed at least one step but had not completed half of the steps by the end of the course
- **Completers:** Learners who completed at least half of the steps or those who bought the certification of participation. This criteria has been set by FutureLearn. According to FutureLearn (https://about.futurelearn.com/research-insights/learners-learning-know), people who completed more than 50% are active users and classified as Completed Learner. People who completed over 90% are qualified to certificate are also classified as Completed Learners by FutureLearn. Since the number of certificate bought students is very low, we have also merged these two groups all together as Completers in our study for the sake of the more accurate algorithm performance.

According to this classification, the number of learners in each group is shown in Table 3. There are clear differences between the behaviours of EPL learners and others. The first difference is that the percentage of EPL completer learners was highest and the percentage of EPL dropouts was lowest overall. We can infer from that, while EPL learners are more likely to complete the course, members of other language-based groups are more likely to dropout. Another difference amongst EPL learners is that, the number of slow paced learners and completing learners were nearly equal. However, in the other groups, the percentage of slow paced learners was far greater than completers. In each group, the smallest percentage of learners were in the completers category.

Table 3. The number of people in each class of course completion before data balancing.

	EPL	EOL	ESL	Overall
Dropouts	316 (49.14%)	237 (54.61%)	1198 (53.77%)	1751 (52.98%)
Slow paced learners	165 (25.66%)	140 (32.26%)	645 (28.95%)	950 (28.74%)
Completers	162 (25.19%)	57 (13.13%)	385 (17.28%)	604 (18.28%)
Total	*643*	*434*	*2228*	*3305*

The number of learners who completed the course is the lowest for each language-based group and the number of learners who dropped out the course is the highest in each language-based group (Table 3). However the volume of the data across the three categories varies extensively.

In order to produce realistic results from the prediction models, we needed to balance the data. To balance the data, a random under sampling method was used. This method enables us to avoid the need to copy large amounts of instances to balance the data which might be misleading for the algorithms. Learners in each category (Drop-outs, Slow Paced Learners, Completers) were randomly selected for prediction by taking the smallest category as reference.

As the smallest amount is 604 for the Completers (Table 3), we have randomly selected 604 learners per the categories of course performance (Dropouts, Slow paced, Completers). Therefore we randomly selected 162 EPL participants, 57 EOL participants and 385 ESL participants per Dropouts, Slow paced and Completers categories. In total, 1812 learners were selected for the experiments.

6.3. Implementation and Prediction Results

The analyses of participants' behaviours and performance could be useful for us to predict their future performance. In this section, we have used the features that are extracted from the learners' engagement in the course (presented in Section 5) for prediction of course completion.

The experiments were carried out with the following widely-used classification algorithms:

- Bayesian Regularized Neural Network (BNN)
- Decision Tree (DT)
- k-Nearest Neighbour (KNN)
- Logistic Regression (LR)
- Naive Bayes (NB)
- Random Forest (RF)
- Support Vector Machine (SVM)

For training the data, the 10-fold cross validation method was applied. Table 4 shows the accuracy results for each algorithm.

According to the results, the Random Forest Model performed best when predicting course performance. The Naive Bayes Model was the worst performer, particularly when predicting of overall course performance.

Table 5 shows the precision, recall, and F-measure metric values for the Random Forest Model. These values demonstrate the better performance of Random Forest compared to the other models investigated.

Table 4. The accuracy results for each algorithm.

	Overall	EPL	EOL	ESL
Bayesian Reg. Neural Network (BNN)	0.9277	0.9547	0.9240	0.9359
Decision Tree (DT)	0.9658	0.9671	0.9474	0.9671
k-Nearest Neighbour (KNN)	0.9404	0.9012	0.8889	0.9489
Logistic Regression (LR)	0.9658	0.9671	0.9474	0.9671
Naive Bayes (NB)	0.8289	0.8704	0.8070	0.8433
Random Forest Model (RFM)	0.9691	0.9733	0.9649	0.9766
Support Vector Machine (SVM)	0.9277	0.9547	0.9240	0.9359

Table 5. Precision, Recall, F-Measure results for the Random Forest algorithm.

	Overall	EPL	EOL	ESL
Precision (Positive Predictive Value)	0.9697	0.9737	0.9649	0.9742
Recall (Sensitivity)	0.9680	0.9774	0.9649	0.9757
F-Measure (F-1 Score)	0.9688	0.9755	0.9649	0.9749

Some algorithms performed up to 10% percent better for the ESL participants (Bayesian Regularized Neural Network, k-Nearest Neighbor, Logistic Regression).

We also tried to identify the most important features for prediction. For all algorithms except the Random Forest Model, the most important features are respectively: (i) number of steps opened, (ii) total completed steps belong to Week 2 (W2), and (iii) total completed steps belong to Week 1 (W1). Interestingly, the Random Forest Model which performed best for each group unlike the other algorithms found a different order of importance in each group detailed in Table 6.

Table 6. Three most important predictive features within the Random Forest Model (which unlike other algorithms found total completed Week 1 steps (W1) to be more important than total completed Week 2 steps (W2)).

Overall	EPL	EOL	ESL
1. steps opened	1. total steps completed W1	1. steps opened	1. steps opened
2. total steps completed W1	2. steps opened	2. total steps completed W1	2. total steps completed W1
3. total steps completed W2	3. total steps completed W2	3. total steps completed W2	3. total steps completed W2

Table 7 shows the confusion matrix of the Random Forest Model. We can infer from Table 7 that the algorithm has the biggest error rate in the prediction of slow paced learners for all language-based groups. With overall (country known) and ESL learners' data, the best result of the Random Forest algorithm was for the prediction of completer learners' performance. With EPL and EOL learners' data, the best result of the Random Forest algorithm was for the prediction of dropout learners.

In the Random Forest algorithm, we also found that *the number of steps are opened by the learner* is the strongest predictive feature for EOL and ESL learners. For EPL participants, the strongest predictive feature is *the total steps completed in W1.*

In classification problems, the accuracy of the prediction algorithms can differ by various factors. In our study on predicting course completion on the FutureLearn platform, the Random Forest algorithm has the best accuracy. The reason why the Random Forest is the best at accurately predicting course completion and identifying most important feature could be because it uses a collection of multi-decision processes to classify.

Table 7. Confusion matrix of Random Forest algorithm for each language group.

Overall Prediction	Behind	Completer	Drop-out	EPL Prediction	Behind	Completer	Drop-out
Behind	591	11	34	Behind	158	5	4
Completer	8	592	0	Completer	2	157	0
Drop-out	5	1	570	Drop-out	2	0	158
EOL Prediction	Behind	Completer	Drop-out	ESL Prediction	Behind	Completer	Drop-out
Behind	53	2	2	Behind	376	7	12
Completer	3	55	0	Completer	4	377	1
Drop-out	1	0	55	Drop-out	5	1	372

6.4. Weekly Prediction of Course Completion for Early Intervention

In order to achieve the best study outcomes, we want to be able to detect participants who may need help when pursuing their study in a course. Our findings suggest that early behaviours are the strongest predictors. Similarly, early intervention will be particularly important for a sustainable MOOC. Effective interventions and feedback will, therefore, rely on our ability to identify as early as possible those participants whose behaviours suggest they are least likely to successfully complete the course.

In this study, we ran the same algorithms for each week cumulatively. Figures 16–19 show the weekly prediction accuracy of the algorithms for each language-based group. In most of the cases, the Naive Bayes (NB) algorithm was the worst performer. It remained under 0.75 for the first two weeks (first 2W) which are the most important weeks for an early intervention.
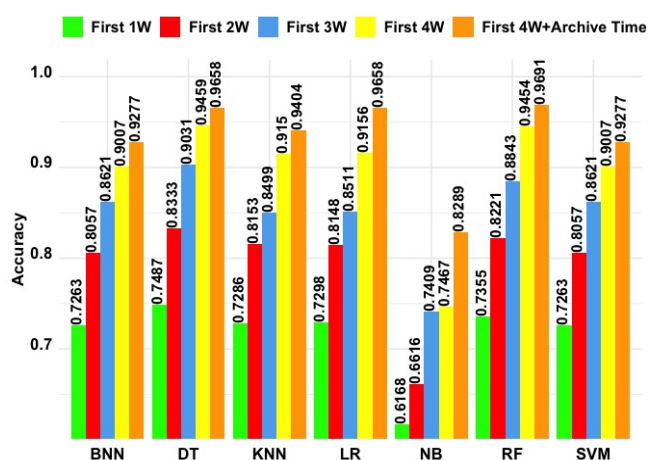
Figure 16. Overall weekly prediction accuracy.

The Random Forest Model worked best for the first four weeks (first 4W) predictions, the Decision Tree model performed slightly better than the Random Forest Model for the first one (first 1W), first two (first 2W) and first three (first 3W) weeks' predictions. The overall performance of the Decision Tree is also very close to the Random Forest Model, the statistical difference between them is insignificant.

The overall weekly performance of the algorithms improved over the weeks. While the accuracy results for the first week predictions are around 75%, it goes up above 80% after the first week except for EPL participants. The accuracy results of ESL participants for the first week remained lower than the other two groups of learners. This suggests it is more difficult to identify early those English as a second language speaking participants who are at risk of leaving the course.

Another distinctive finding of the weekly predictions is the most strongly predictive features. The results show that when the first week has been completed, participants' engagement in conversations is one of the predictive features. For example, the average length of the comment

is the third predictive feature with the Random Forest Model, the number of comment is the third predictive feature with the Decision Tree. In the rest of the course, participants' behaviour in completion of course steps is more stongly predictive.
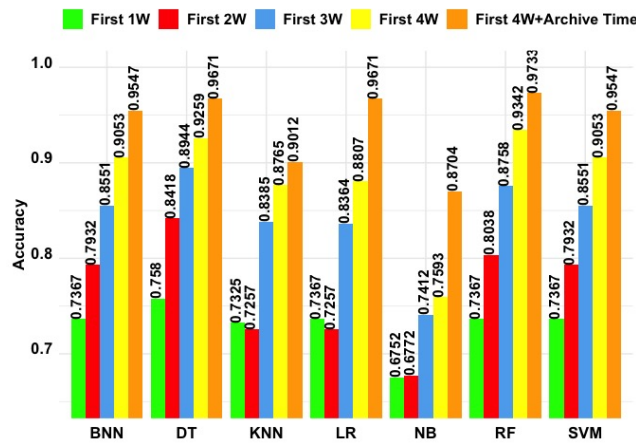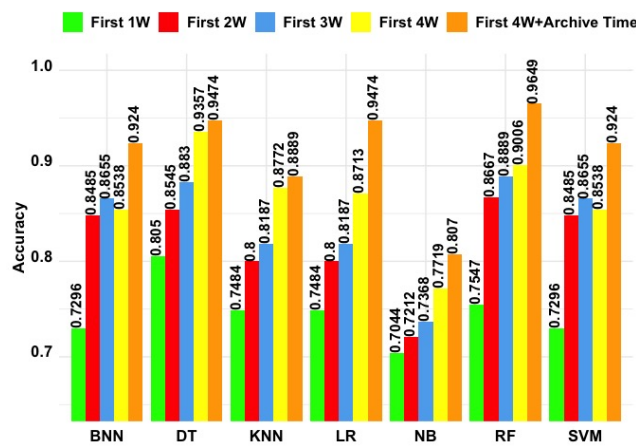

Figure 17. EPL weekly prediction accuracy.
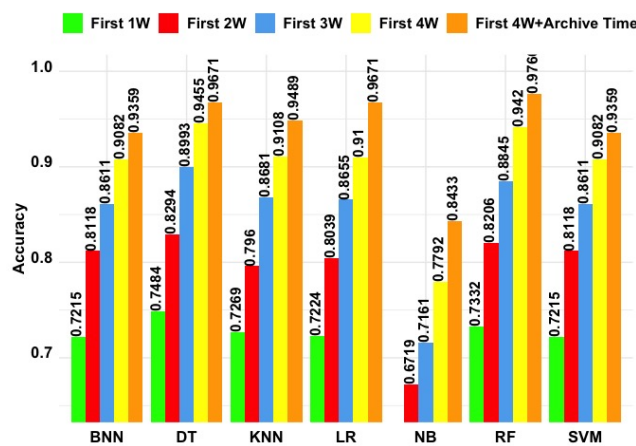

Figure 18. EOL weekly prediction accuracy.

Figure 19. ESL weekly prediction accuracy.

7. Discussion and Conclusions

A more recent study where findings were based on data from international participants in a MOOC focussed on social enterprise education, Calvo et al. [43] specifically identify linguistic and cultural barriers as inhibiting learner's access to MOOCs. Much of the early speculative literature which promotes the potential for MOOCs highlightes the value of free and open education. Subsequent attitudinal and implementation studies reveal barriers to accessibility frequently focusing on cultural and linguistic aspects. Whilst it may be evident that a large proportions of MOOC participants are drawn from developing countries [2] work remains to be done by MOOC providers to enhance the effective usefulness of this growing set of rich educational resources. Finding ways to automatically identify key features associated with learners (such as their approximate linguistic backgrounds) offers a means for MOOC platform providers and course authoring teams to realistically consider broad brush approaches to personalisation. Furthermore this approach could also potentially be used to provide data to enable effective localisation, the need for which has been identified by Castello et al. [44] Our reasoning for focusing on English language competencies was based on the observation that a considerable proportion of MOOCs at present are conducted in the English language, coupled with an understanding that socially active learners (those who are involved participants in online discussion based tasks and exercises) are most likely to complete the course [42].

Our research has analysed the social engagement and course completion performance of participants categorised by first language groupings. We find that participants whose first language is English are able to make more active use of the platform and are most likely to complete the course. This inequality can potentially be addressed if we are able to successfully identify learners with other linguistic markers and provide tailored support or customised interventions to narrow this achievement gap. The research presented in this paper identifies some initial steps that could contribute towards such an approach.

This study set out to find a better way of categorising MOOC learners and their behaviours such that it might be possible in future to automatically or semi-automatically tailor learning material to enhance the chance of success and benefit eventual learning outcomes. Approaches adopted included:

- Analysing comment data using regular expressions to categorise learner's linguistic antecedents.
- Comparing course engagement of learners within different linguistic categories using progress participation and completion as key indicators.
- Investigating the viability of establishing the use of a prediction model based on these data.

We compared the course engagement of learners grouped according to whether or not their first language is English. The first challenge in this research was to automatically detect the participants who speak English as a second language, which is one of the research questions presented in Section 3. Regular expressions have been used in MOOCs to extract hashtags and keywords from text [45,46]. Our study proposed a novel method using regular expressions to identify someone's first language from their comment in discussions. It is observed that our proposed method with regex enables us to accurately identify more numbers of learners than we directly identified from the course survey answers.

Differently from the existing literature, our study analysed the distinguish differences among the behaviours of participants diverse in first language. We especially analysed and compared their behaviours in completing course steps, attendance in forum conversations, and following each other to answer the second research question in Section 3. We found that whilst learners in the different categories sometimes show similar behaviours, there are also ways in their behaviours differ.

The overall participation in the course supports the findings from previous studies [40] as the participation steadily decreased over the weeks and completion of the course remained low. However, our findings showed that there is a difference in participation and completion of the course among the language-based groups. For example, while most of the people whose first language was not English

(EOL and ESL) completed none of the course steps, those whose first language was English (EPL) showed better performance in completion steps.

The proportion of participants attending discussion forums are very similar which is around 30%.

While the majority of learners in each group posted a very small number of comments, it is seen that the outliers who performed better are usually from the learners whom English is their first language.

The biggest difference among behaviours is observed in being followed. The participants whose their first language is English are followed by others far more than those in two other groups. The reason needs to be investigated though, it can be intuitively said that the clarity of written language in comments posted by the learners whom English was their first language might lead others to follow them.

A further research question (Section 3) is whether we can use the differences in behaviours of categorised participants by their first language to build a predictive model and predict the participants' completion of course.

The observed behaviours are extracted as features for prediction models. When we consider those features which best predict the course performance of participants, the Random Forest Model gave the highest accuracy across the seven prediction algorithms showed. In the Random Forest Model whilst the total completed steps belong to Week 1 was the strongest predictive feature for the learners whom English was their first language, the number of opened steps was the strongest predictor for the learners whom English was their second language. In the other six models the number of steps opened by a learner was the strongest predictive feature across the all learner groups. Finding the behaviours in the first week correlated to completion is previously confirmed by the study of Jiang et al. [47]. Differently, their study shows that social engagements in Week 1 is strongly correlated while the findings of our study shows that the engagement with the course steps in Week 1 is the most predictive.

The results of this research can be concluded as:

- Regular expression patterns are useful to automate the process of identifying English as a second language participants in MOOCs as long as the participants mention about their first language, nationality, city or country they are from.
- The participants whose first language is English are usually more active in engaging with the course than the others. For example, they post more numbers of comments to the discussions, they write longer comments, and they are more likely to be followed by others.
- The participants whose first language is English are more likely to complete more numbers of course steps than others.
- The Random Forest algorithm performed best for prediction of the course completion. The Random Forest algorithm also performed best for the weekly prediction.
- The total steps completed in the first week is most predictive feature for the The participants whose first language is English while the number of steps opened is the most predictive for the learners whom English was not their first language (EOL and ESL learners). However, the top three predictive features are the same for all categories.

8. Future Work

In this work, to make more accurate identification of English as a second language participants, we have proposed a novel method with regular expression patterns by using the comments posted to the discussions. We have investigated and identified the sentences where the participants explicitly said where they are from, what their nation is or what their first language is. Based upon these data, we have created the regular expression patterns to identify languages.

However, there might be other patterns where people express themselves in different sentence structures in other courses. We need a more accurate automatic system for detection of English as a second language participants in MOOCs.

Another limitation of the identification of English as a second language participants with regular expressions is that it may not be very extendable to other MOOCs unless they encourage their participants to talk about their first languages. The applicability of the method to other MOOCs needs to be investigated.

This study was undertaken within a single MOOC (the fourth FutureLearn *Understanding Language: Learning and Teaching*). It would be particularly interesting to replicate the approach (1) with larger data sets from successive run of the same MOOC, (2) across a range of disciplines. For example, we would like to know whether successful behaviours in different discipline groups conform to the same predictive patterns that we have identified in this study.

A further extension to this work would be to identify and adopt an approach which would enable us to identify learners' language category more accurately. To accomplish this, we will work on methods to automatically detect differences in language fluency for the learners whom English was not their first language (EOL and ESL learners).

Additionally, we plan to complete more in-depth analysis of language-based groups' behaviour and performance. This may help us identify suggestions of optimal study patterns for learners according to their language history. This evidence-based approach might help learners improve their study experience within MOOCs.

Apart from these future research directions, the findings from our research could be used by the MOOC providers (authors and platform creators) for re-designing their courses and platforms where English as a second language speakers would be most likely to benefit. The findings show that the participants whose first language is English are more likely to complete a greater number of steps and they are more actively engaged in the discussions. Preparing a MOOC and a platform which is encouraging for English as a second language speakers to more actively attend to discussions may cause for them to complete a greater number of the steps.

References

1. Almutairi, F. The Impact of Integrating MOOCs into Campus Courses on Student Engagement. Ph.D. Thesis, University of Southampton, Southampton, UK, 2018.
2. Liyanagunawardena, T.R.; Williams, S.; Adams, A.A. *The Impact and Reach of MOOCs: A Developing Countries' Perspective*; eLearning Papers; University of Reading: Reading, UK, 2014; pp. 38–46. Available online: http://centaur.reading.ac.uk/32452/1/In-depth_33_1.pdf (accessed on 14 May 2019).
3. Dillahunt, T.R.; Wang, B.Z.; Teasley, S. Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *Int. Rev. Res. Open Distrib. Learn.* **2014**, *15*. [CrossRef]
4. Mamgain, N.; Sharma, A.; Goyal, P. Learner's perspective on video-viewing features offered by MOOC providers: Coursera and edX. In Proceedings of the MOOC 2014 IEEE International Conference on Innovation and Technology in Education (MITE), Patiala, India, 19–20 December 2014; pp. 331–336.

5. Eriksson, T.; Adawi, T.; Stöhr, C. "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *J. Comput. High. Educ.* **2017**, *29*, 133–146. [CrossRef]

6. Cho, M.H.; Byun, M.K. Nonnative English-Speaking Students' Lived Learning Experiences with MOOCs in a Regular College Classroom. *Int. Rev. Res. Open Distrib. Learn.* **2017**, *18*, 173–190. [CrossRef]

7. Kizilcec, R.F.; Saltarelli, A.J.; Reich, J.; Cohen, G.L. Closing global achievement gaps in MOOCs. *Science* **2017**, *355*, 251–252. [CrossRef] [PubMed]

8. Uchidiuno, J.; Ogan, A.; Yarzebinski, E.; Hammer, J. Understanding ESL Students' Motivations to Increase MOOC Accessibility. In Proceedings of the Third (2016) ACM conference on Learning @ Scale, Edinburgh, UK, 25–26 April 2016; pp. 169–172.

9. Colas, J.F.; Sloep, P.B.; Garreta-Domingo, M. The effect of multilingual facilitation on active participation in MOOCs. *Int. Rev. Res. Open Distrib. Learn.* **2016**, *17*. [CrossRef]

10. Chatti, M.A.; Lukarov, V.; Thüs, H.; Muslim, A.; Yousef, A.M.F.; Wahid, U.; Greven, C.; Chakrabarti, A.; Schroeder, U. Learning analytics: Challenges and future research directions. *Eleed* **2014**, *10*. Available online: https://eleed.campussource.de/archive/10/4035 (accessed on 14 May 2019).

11. Ramesh, A.; Goldwasser, D.; Huang, B.; Daume III, H.; Getoor, L. Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In Proceedings of the First ACM Conference on Learning @ Scale Conference, Atlanta, GA, USA, 4–5 March 2014; pp. 157–158.

12. Papadakis, S.; Kalogiannakis, M.; Sifaki, E.; Vidakis, N. Access moodle using smart mobile phones. A case study in a Greek University. In *Interactivity, Game Creation, Design, Learning, and Innovation*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 376–385.

13. Klüsener, M.; Fortenbacher, A. Predicting students' success based on forum activities in MOOCs. In Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Warsaw, Poland, 24–26 September 2015; Volume 2, pp. 925–928.

14. 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 Trans. Learn. Technol.* **2014**, *7*, 346–359. [CrossRef]

15. DeBoer, J.; Breslow, L. Tracking progress: Predictors of students' weekly achievement during a circuits and electronics MOOC. In Proceedings of the First ACM Conference on Learning @ Scale Conference, Atlanta, GA, USA, 4–5 March 2014; pp. 169–170.

16. Kennedy, G.; Coffrin, C.; De Barba, P.; Corrin, L. Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Poughkeepsie, NY, USA, 16–20 March 2015; pp. 136–140.

17. Xu, B.; Yang, D. Motivation classification and grade prediction for MOOCs learners. *Comput. Intell. Neurosci.* **2016**, *2016*, 2174613. [CrossRef] [PubMed]

18. Ye, C.; Biswas, G. Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *J. Learn. Anal.* **2014**, *1*, 169–172. [CrossRef]

19. Yang, T.Y.; Brinton, C.G.; Joe-Wong, C.; Chiang, M. Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 716–728. [CrossRef]

20. Brouns, F.; Serrano Martínez-Santos, N.; Civera, J.; Kalz, M.; Juan, A. Supporting language diversity of European MOOCs with the EMMA platform. In Proceedings of the Third European MOOCs Stakeholder Summit (eMOOCs2015), Mons, Belgium, 18–20 May 2015; Lebrun, M., Ebner, M., Eds.; 2015; pp. 157–165. Available online: http://dspace.ou.nl/handle/1820/6026 (accessed on 14 May 2019).

21. Wu, W.; Bai, Q. Why Do the MOOC Learners Drop Out of the School?—Based on the Investigation of MOOC Learners on Some Chinese MOOC Platforms. In Proceedings of the 2018 1st International Cognitive Cities Conference (IC3), Okinawa, Japan, 7–9 August 2018; pp. 299–304.

22. Wang, Q.; Chen, B.; Fan, Y.; Zhang, G. MOOCs as an Alternative for Teacher Professional Development: Examining Learner Persistence in One Chinese MOOC. 2018. Available online: http://dl4d.org/wp-content/uploads/2018/05/China-MOOC.pdf (accessed on 14 May 2019).

23. Beaven, T.; Codreanu, T.; Creuzé, A. Motivation in a language MOOC: Issues for course designers. In *Language MOOCs: Providing Learning, Transcending Boundaries*; De Gruyter Open: Berlin, Germany, 2014; pp. 48–66.

24. Türkay, S.; Eidelman, H.; Rosen, Y.; Seaton, D.; Lopez, G.; Whitehill, J. Getting to Know English Language Learners in MOOCs: Their Motivations, Behaviors, and Outcomes. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, Cambridge, MA, USA, 20–21 April 2017; pp. 209–212.

25. Uchidiuno, J.O.; Ogan, A.; Yarzebinski, E.; Hammer, J. Going Global: Understanding English Language Learners' Student Motivation in English-Language MOOCs. *Int. J. Artif. Intell. Educ.* **2017**, *28*, 528–552. [CrossRef]

26. Uchidiuno, J.; Koedinger, K.; Hammer, J.; Yarzebinski, E.; Ogan, A. How Do English Language Learners Interact with Different Content Types in MOOC Videos? *Int. J. Artif. Intell. Educ.* **2018**, *28*, 508–527. [CrossRef]

27. Obari, H.; Lambacher, S. Impact of a Blended Environment with m-Learning on EFL Skills. In Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands, 20–23 August 2014; pp. 267–272. [CrossRef]

28. Nie, T.; Hu, J. EFL Students' Satisfaction with the College English Education in the MOOC: An Empirical Study. In Proceedings of the 2018 2nd International Conference on Education, Economics and Management Research (ICEEMR 2018), Singapore, 9–10 June 2018.

29. Fuchs, C. The Structural and Dialogic Aspects of Language Massive Open Online Courses (LMOOCs): A Case Study. In *Computer-Assisted Language Learning: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2019; pp. 1540–1562.

30. Martín-Monje, E.; Castrillo, M.D.; Mañana-Rodríguez, J. Understanding online interaction in language MOOCs through learning analytics. *Comput. Assist. Lang. Learn.* **2018**, *31*, 251–272. [CrossRef]

31. Qian, K.; Bax, S. *Beyond the Language Classroom: Researching MOOCs and Other Innovations*; Research-Publishing: Voillans, France, 2017; pp. 15–27. [CrossRef]

32. Tan, Y.; Zhang, X.; Luo, H.; Sun, Y.; Xu, S. Learning Profiles, Behaviors and Outcomes: Investigating International Students' Learning Experience in an English MOOC. In Proceedings of the 2018 International Symposium on Educational Technology (ISET), Osaka, Japan, 31 July–2 August 2018; pp. 214–218.

33. Hemavathy, R.; Harshini, S. Adaptive Learning in Computing for Non-English Speakers. *J. Comput. Sci. Syst. Biol.* **2017**, *10*, 61–63.

34. Guo, P.J. Non-native English speakers learning computer programming: Barriers, desires, and design opportunities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 396.

35. Rimbaud, Y.; McEwan, T.; Lawson, A.; Cairncross, S. Adaptive Learning in Computing for non-native Speakers. In Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE), Madrid, Spain, 22–25 October 2014; pp. 1–4.

36. De Waard, I.; Demeulenaere, K. The MOOC-CLIL project: Using MOOCs to increase language, and social and online learning skills for 5th grade K-12 students. In *Beyond the Language Classroom: Researching MOOCs and Other Innovations*; Research-Publishing: Voillans, France, 2017; pp. 29–42. [CrossRef]

37. Reilly, E.D.; Williams, K.M.; Stafford, R.E.; Corliss, S.B.; Walkow, J.C.; Kidwell, D.K. Global Times Call for Global Measures: Investigating Automated Essay Scoring in Linguistically-Diverse MOOCs. *Online Learn.* **2016**, *20*, 217–229. [CrossRef]

38. Duru, I.; Sunar, A.S.; Dogan, G.; White, S. Challenges of Identifying Second Language English Speakers in MOOCs. *European Summit on Massive Open Online Courses*; Springer: Cham, Switzerland, 2017; pp. 188–196.

39. Uchidiuno, J.; Ogan, A.; Koedinger, K.R.; Yarzebinski, E.; Hammer, J. Browser language preferences as a metric for identifying ESL speakers in MOOCs. In Proceedings of the 3rd ACM conference on Learning @ Scale, Edinburgh, UK, 25–26 April 2016; pp. 277–280. [CrossRef]

40. Clow, D. MOOCs and the funnel of participation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium, 8–13 April 2013; pp. 185–189.

41. Duru, I.; Sunar, A.S.; Dogan, G.; Diri, B. Investigation of social contributions of language oriented MOOC learner groups. In Proceedings of the 2017 International Conference on IEEE Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017; pp. 667–670. [CrossRef]

42. Sunar, A.S.; White, S.; Abdullah, N.A.; Davis, H.C. How learners' interactions sustain engagement: A MOOC case study. *IEEE Trans. Learn. Technol.* **2017**, *10*, 475–487. [CrossRef]

43. Calvo, S.; Morales, A.; Wade, J. The use of MOOCs in social enterprise education: An evaluation of a North–South collaborative FutureLearn program. *J. Small Bus. Entrep.* **2019**, *31*, 201–223. [CrossRef]

44. Castillo, N.M.; Lee, J.; Zahra, F.T.; Wagner, D.A. MOOCS for development: Trends, challenges, and opportunities. *Int. Technol. Int. Dev.* **2015**, *11*, 35.

45. Acosta, E.S.; Otero, J.J.E. Automated assessment of free text questions for MOOC using regular expressions. *Inf. Resour. Manag. J. (IRMJ)* **2014**, *27*, 1–13. [CrossRef]

46. An, Y.H.; Chandresekaran, M.K.; Kan, M.Y.; Fu, Y. The MUIR Framework: Cross-Linking MOOC Resources to Enhance Discussion Forums. In Proceedings of the International Conference on Theory and Practice of Digital Libraries, Porto, Portugal, 10–13 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 208–219.

47. Jiang, S.; Williams, A.; Schenke, K.; Warschauer, M.; O'Dowd, D. Predicting MOOC performance with Week 1 behavior. In Proceedings of the 7th International Conference on Educational Data Mining 2014, London, UK, 4–7 July 2014.