Slow and Steady or Fast and Furious: An Analysis of Completion Duration in open.uom.lk

Lakmini Bandarigodage, Tiloka de Silva, Eshana Ranasinghe and Vishaka Nanayakkara
University of Moratuwa, Sri Lanka

bandarigodageln.20@uom.lk
tilokad@uom.lk
enranasinghe1@gmail.com
vishaka@cse.mrt.ac.lk

Abstract: The open.uom.lk platform is an open learning platform developed by the University of Moratuwa, Sri Lanka, offering free asynchronous online IT related courses. The platform was developed as an initiative to meet the growing human capital requirement of the IT industry in Sri Lanka, one of the only sectors showing strong growth even during the country's recent economic crisis. In the two years since the release of its flagship course Trainee Full-Stack Developer, there have been over 270,000 enrolled learners. The course consists of six subjects covering Python programming, web design, and a professional practice module that includes a capstone project. Despite the massive number of enrollments, as in many other MOOCs, completion rates remain low. For instance, for the two beginner courses in programming and web design, into which all participants are enrolled automatically, completion rates stand at 7-10%. Completion rates for subsequent subjects are higher given that enrollment is conditional on completing prerequisites. From a planning perspective, it is not just estimates of how many participants complete the programs but also the time they take to do so that are integral for determining the future growth of the IT industry. As such, this study aims to develop a model for course completion using survival analysis which has the advantage of being able to model censored data (e.g. when duration is unknown due to non-completion) and provide forecasts of completion rates at different points in time. The analysis uses student activity completion reports and demographic information of participants including employment and educational background. The findings suggest that the course interaction variables, particularly the speed of completing early activities, are more important for predicting course completion than the demographic variables, though education and employment status are also significantly correlated with completion. The survival model could then be used for purposes of predicting both the completion probability as well as the timeframe within which the completion may be expected.

Keywords: MOOCs, Course completion, Completion duration, Survival Analysis, Learning analytics

1. Introduction

Massive Open Online Courses (MOOCs) have become increasingly widespread since their inception in the 2010s and are now an important part of the education landscape. The disruptions caused by the Covid-19 pandemic including social distancing measures and closure of physical learning spaces provided an opportunity for MOOCs to be positioned as a viable solution for the education sector (Rulinawaty et al, 2023). Given the massive uptake and potential for outreach, there is an active debate on their role and value in higher education: while online, asynchronous learning allows a large amount of independence and flexibility for learners as well as coverage of a wide array of subjects for larger student numbers, these courses face challenges of low persistence and completion rates (Impey and Formanek, 2021).

The Centre for Open and Distance Learning (CODL) of the University of Moratuwa, Sri Lanka, launched the open.uom.lk platform for asynchronous online learning in 2021 as a response to the shortage of skilled workers in the IT industry in the country. Despite the digital technology sector showing strong economic growth even during the country's recent economic crisis, a clear gap between the market demand and supply of ICT skills has been identified (ICTA, 2019). Accordingly, the Trainee Full-Stack Developer (TFSD) course was developed in consultation with members of the IT industry to cover the knowledge and skills required by the sector (Ranasinghe et al., 2023). The course is offered completely free of charge and is open to anyone.

The TFSD course consists of 6 subjects: two courses in Python programming, three courses in web development, and the final course that covers professional practice and a capstone project. A registering student is automatically enrolled into the introductory course in Python programming and web development. Enrolment into subsequent courses is based on the completion of the prerequisites. Given that the goal of the program involves the creation of skilled workers who are employable in the IT industry, the course is rigorous. As such, (like in many other MOOCs) completion rates for the course are low. While 273,179 users have enrolled in the program as of the 21st of May 2024, only 350 of the enrolled users had completed the full course, while only 19,316 and 13,291 had completed the introductory courses for programming and web development, respectively.

With the observed high levels of enrolment and low completion rates, and the urgent requirement for skilled workers for the IT industry, it would be useful to predict not just how many students will complete but also the time horizon within which those numbers can be expected. While standard classification models do the former, they do not necessarily make use of the duration of completion and therefore cannot be used for the latter. Survival analysis provides an alternative for this type of scenario given that it makes predictions on the time to an event, rather than the event itself.

This paper aims to use survival analysis to model student completion in the open.uom.lk platform, making use of student activity reports from the Web Design for Beginners course and demographic information. The development of the model would facilitate the prediction of completions within a particular time frame, while also allowing for the identification of key factors that contribute to the timely completion of the course. The former is important from an overall sector planning perspective whereas the latter is useful for further developing the course itself.

The growing use of learning management systems (LMS) and educational technology has resulted in the creation of highly detailed educational datasets, facilitating the development of learning analytics for improving student retention and reducing student attrition (Villano et al., 2018). MOOCs provide a particularly rich environment for this type of analysis given the large user bases and digital format of the courses and the widespread problem of low completion. Previous work on predicting dropouts or completion has identified student-specific demographic and social factors as well as interaction-specific factors that can be used to predict various outcomes including course completion, probability of dropping out, prediction of scores, etc. Moreno-Marcos et al (2018) provide a system review of this literature.

As discussed in Moreno-Marcos et al. (2018), a majority of publications focused on predictive models using MOOCs make use of variables that can be constructed using the activity logs and records generated within the LMS itself. These range from simple measures such as time spent on the LMS, total clicks, and frequency of access (Ulfa and Fatawi, 2021; Damianov et al, 2009; Vengross and Bourbeau, 2006), to more sophisticated measures such as logs on video events such as playing, pausing and skipping (Brinton et al. 2016) or time spent on specific activities such as solving exercises or watching videos (Ruiperez-Valiente at al. 2018; Boyer et al. 2015). Many studies also make use of demographic characteristics obtained either from course registration forms or surveys carried out on course participants. Commonly used characteristics include age, gender and educational background, including familiarity with the course topic (e.g. Greene et al. 2015; Al-Shabandar et al. 2017; Labrador et al. 2019).

There is a large literature making use of classification models including logistic regression, support vector machines, decision trees, and random forests (among many others) for predicting completions or drop-outs in MOOCs. For instance, a study by Chaplot et al. (2015) finds that by using Sentiment Analysis and Neural Networks they could predict the students who are likely to drop out. Their method had high accuracy (88%) and low occurrences of false negatives (0.095). Machine learning can also be used to predict student attrition such as in the study by Matz et al. (2023) who used user engagement metrics and socio-demographic characteristics data for their linear and non-learning machine learning model. They found their models had a 78% average accuracy. Machine learning and deep learning techniques have been widely applied when predicting student outcomes in online courses and evident in the review by Alhothali et al. (2022).

Work on the application of survival analysis models to this problem is less extensive. However, as argued by Greene et al. (2015) survival analysis has the advantage of using information on the timing of events (in this case, completion), while also accounting for censored data (i.e. participants for whom the event is not experienced over the course of the analysis but may do so in the future) which gives rise to better estimates of the effect of covariates used to predict event probabilities. Work by Green et al. (2015), Gitinabard et al. (2018), Labrador et al. (2019), Wintermute et al. (2021) provide notable examples of survival analysis applications to the learning analytics context. Given this background, this research applies survival analysis techniques to model course completion in the open.uom.lk platform, making use of both demographic and interaction-specific characteristics. In doing so, it not only provides actionable insights to local policymakers and administrators of the open learning platform but also adds to the literature on the potential for open and distance learning of IT-related subjects in a developing country context.

2. Data

This paper makes use of activity completion reports of Web Design for Beginners, the introductory web development course in the TFSD program offered on the open.uom.lk platform. The Web Design course has a total of 45 activities that must be completed sequentially before the overall course can be completed. The activity completion report has information about the date and the time of the completion of each activity for each user, allowing for the construction of indicators for activity and course completion at the time of report generation as well as the duration taken for completion of individual activities as well as the overall course.

The activity completion data is supplemented with demographic data from the registration survey. The available demographic information includes gender, level, field of education, and field of employment. Given that the registration survey was introduced after the Web Design course was launched the demographic data is only available for a subset of 25,625 users. Moreover, since Web Design for Beginners is one of the introductory courses in the program, all students who register for the TFSD course are automatically enrolled. Accordingly, a large number of users who appear in the activity report have not completed a single activity. Given the issue of class imbalance with non-completions greatly outnumbering completions, we restrict the analysis to those who have completed at least the first activity, resulting in a sample of 5,802 observations.

The percentage of dropouts and median duration for each activity in the course highlights several bottlenecks. For instance, the dropout rates are highest after the 1st and 2nd activities at 15% and 16%, respectively, while the next bottleneck is at activity 6 (10% dropout). We use this information to create course interaction variables to be included in the model. Specifically, we construct indicators for completion of the first five and ten activities, completion of the 2nd activity within 24 hours of the first, completion of the first five activities within three days of the first activity and the completion of the first ten activities within seven days of the first. The last three variables are constructed on the basis that behaviour of the learner in the first week of starting the course can predict student performance and course completion (Jiang et al. 2014). Finally, the response variables are an indicator for completion of the Web Design course and time taken to complete the course (right censored as many students had not completed the course at the time of report generation).

The descriptive statistics show that among those who have completed at least the first activity in the course, the completion rate is 28%, with a mean time to complete of 51 days. The median time to completion was 13 days. The fastest time to completion was 36 minutes while the longest time to complete was 681 days. Figure 1b illustrates the distribution of the completion duration for those who completed the course.

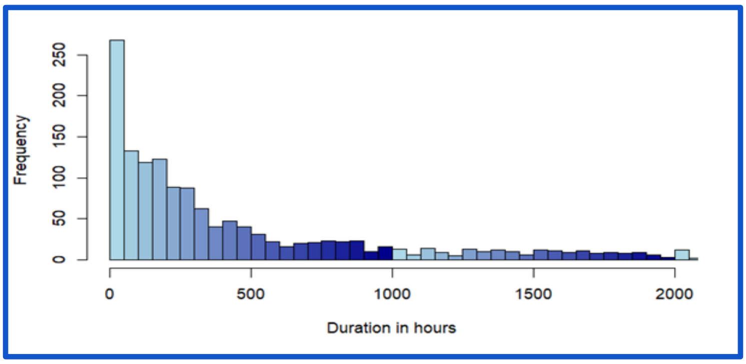

Figure 1: Histogram of time taken to complete the course (among completing students)

Table 1 provides the list of variables used and their descriptive statistics for the sample used for the analysis.

Table 1: Descriptive statistics

Variable	Mean
Completed whole course	0.283
Time taken to complete (only for those who completed) (days)	51.22
Completed 5 activities	0.608
Completed 10 activities	0.290
Completed the 2nd activity within 24 hours of 1st activity	0.673

Variable	Mean
Completed 5 activities within 3 days of 1st activity	0.399
Completed 10 activities within 1 week of 1st activity	0.290
Gender (excluded category = male)	
Female	0.374
Secondary education status (excluded category = Have not done Ordinary Level exam)	
Completed Ordinary Level exam but not passed Advanced Level exam	0.050
Completed Ordinary level exam but not sat for Advanced Level exam yet	0.146
Sat for Advanced Level exam and waiting for results	0.060
Completed Advanced Level exam	0.543
Higher education (excluded category = Have not done a degree)	
Doing a degree in the IT field	0.320
Doing a degree in a non-IT field	0.084
Completed the degree in IT field	0.033
Completed the degree in a non-IT field	0.045
Employment (excluded category = not working)	
Working in a non-IT company	0.089
Working in an IT company	0.046
Self-employed	0.055

3. Methods

This paper uses survival analysis to predict completion and time to completion of the Web Design for Beginners course. The activity logs allow for the computation of the duration taken to complete in minutes, with the origin time period (t=0) considered to be completion of the 1st activity in the course. Given that the event in this study is completion of the course, survival in this model refers to non-completion as of the date of the activity report being generated.

We start by plotting Kaplan-Meier curves for the selected categorical covariates. The Kaplan Meier curve graphically represents probability of surviving in a given length of time (Goel, et al. 2010), allowing for a visual representation of the probability of not completing the course over time. By plotting the curves for the selected covariates, we can visualize differences in survival functions between groups, if any. The p-value of the logrank test for comparing the survival curves of the groups is also reported, where rejection of the null hypothesis suggests significant differences in survival rates between groups. These results form the basis for the selection of covariates for the multivariate analysis that follows.

A Cox proportional hazards model is then fitted to the data, to estimate the level of risk associated with the failure event (in this case, completion) for the selected covariates. The model is similar to a logistic regression for course completion with the key difference that the Cox model takes into account the time unit the event occurs. Accordingly, the Cox model has the advantage of being able to predict probabilities of completion over different time horizons (van der Net et al., 2008). Three separate specifications of the Cox model are estimated: the first uses only demographic variables as covariates, the second only the course interaction variables and the third uses both sets of variables.

The concordance statistic, which is commonly used for evaluating goodness of fit for Cox models, is reported for the three models. The concordance statistic is the fraction of all possible pairs of predicted and actual values that are concordant (i.e. have the same ordering) (Therneau and Watson, 2017) - a higher concordance statistic refers to higher level of agreement between and actual values.

4. Results

Figure 2 plots the Kaplan-Meier curve for the Web Design course completion. Note that survival here refers to not completing the course. As the Graph shows, the probability of not completing the course at time zero is 1, with the probability of non-completion declining thereafter. For instance, the probability of taking more than 50 days to complete the course is approximately 0.75. However, the curve plateaus out - the probability of taking more than 300 days to complete the course remains above 0.6. The curve is steepest in the first 20-30 days, indicating that most completions occur around this duration. These features are consistent with the low completion rates that are observed in the data and the relatively shorter completion times for those who complete the course.

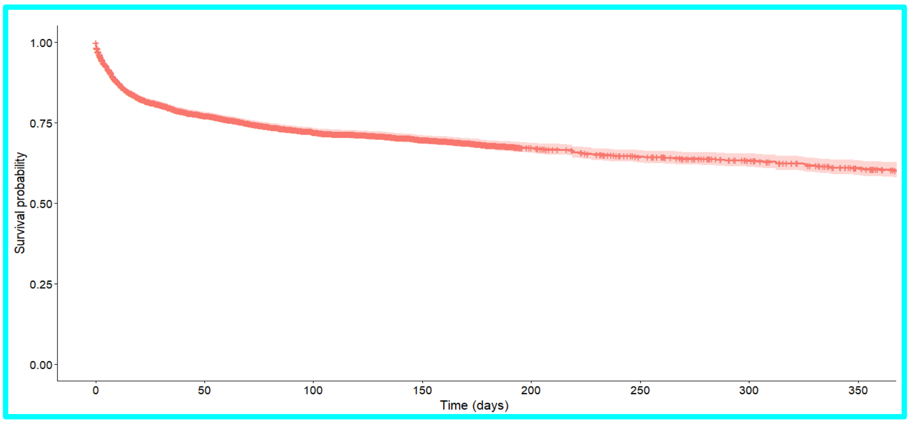

Figure 2: Kaplan-Meier plot for the probability of completing the course

Figure 3 plots the Kaplan-Meier curves for some of the covariates considered in the analysis, both demographic and interaction specific variables. The p-values of the logrank tests comparing survival rates between groups is less than 5% for all covariates considered and smaller than 0.0001 for all covariates other than gender, suggesting that there are significant differences between the survival rates of these groups. For instance, the survival rate for males has a steeper fall than that of females, suggesting that male students are more likely to complete the course. Similarly, students following a degree in an IT-related field are more likely to complete the course than those who are not, students who are not working are more likely to complete than those working in a non-IT company and so on. It should be noted that differences between groups are strongest when considering the variables related to course interactions. For example, the probability that a student who has completed the first ten activities within one week will take more than 20 days to complete the course is less than 0.5, whereas the same probability for a student who did not complete the same activities within a week is close to 0.9. These graphs motivate the use of the selected covariates in the Cox models that follow.

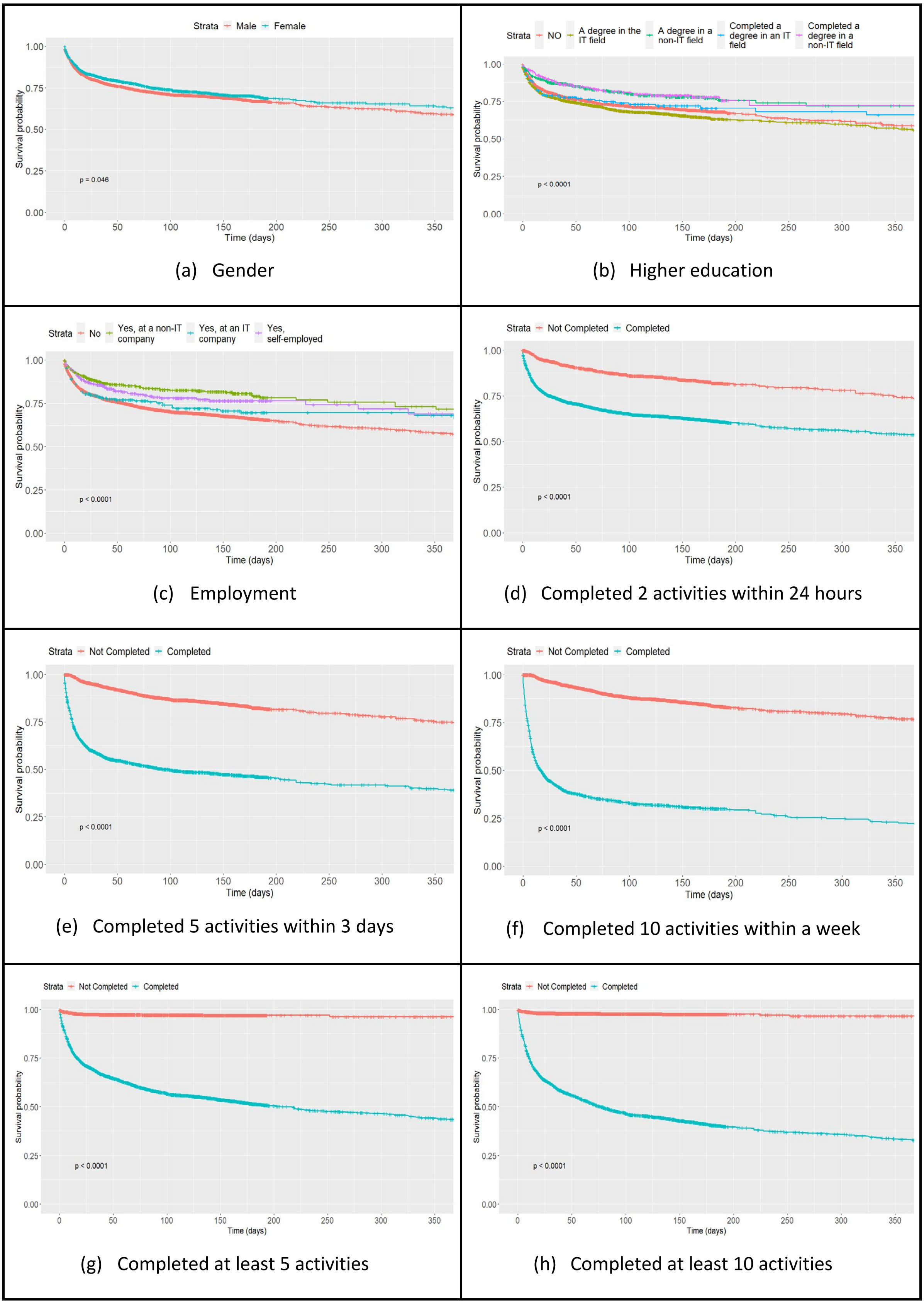

Figure 3: Kaplan-Meier plot for covariates

The output of the Cox regressions for three sets of covariates (only demographic, only course interactions, and both demographic and course interactions) are presented in Table 2.

Table 2: Estimates of the Cox proportional hazard models for course completion

Explanatory variables	**(1)**	**(2)**	**(3)**
Female	-0.123 * [0.032]		0.010 [0.865]
Secondary education (reference: have not done the O/L exam)			
Completed OLs not passed ALS	-0.265 [0.056]		-0.345* [0.014]
Completed OLs not done ALs	-0.203* [0.036]		-0.397*** [0.000]
Sat for ALs	-0.292* [0.029]		-0.335* [0.013]
Completed ALs	-0.281** [0.003]		-0.567*** [0.000]
Employment status (reference: not working)			
Non-IT company	-0.390*** [0.00]		-0.234* [0.048]
An IT company	-0.054 [0.676]		0.106 [0.406]
Self-employed	-0.378** [0.006]		-0.386** [0.006]
Higher education (reference: have not done/not doing a degree)			
Doing a degree in the IT field	0.305*** [0.000]		0.108 [0.127]
Doing a degree in a non-IT field	-0.250* [0.043]		-0.058 [0.640]
Completed degree in IT field	0.167 [0.300]		0.323* [0.044]
Course interactions			
Completed degree in a non-IT field	-0.162 [0.387]		-0.211 [0.264]
Completed up to 5 activities		-1.897** [0.001]	-1.884** [0.002]
Completed up to 10 activities		4.370 *** [0.000]	4.359*** [0.000]
Completed first and second activities within 24 hours		-0.156 [0.055]	-0.151 [0.061]
Completed first five activities within 3 days		0.170* [0.041]	0.195* [0.02]
Completed first ten activities within 1 week		1.156 *** [0.000]	1.171*** [0.000]
Model concordance	*0.5689*	*0.8429*	*0.8209*

Note: The table reports the coefficients and associated p-values (in brackets) from the estimated Cox models. Significance levels: * 0.05 ** 0.01 *** 0.001

The table reports the estimated coefficients and p-values from the three regressions. Focusing on the model with only demographic covariates (Column 1) first, gender, secondary education status, employment status and higher education status all have significant effects on the probability of completing the course. Negative coefficients show negative associations so female students are less likely to complete than male students; students who have completed the Advanced Level or Ordinary Level examination are less likely to complete than students who have not yet completed the Ordinary Level examination; students who are not working are more likely to complete than those who are; and students pursuing a degree in an IT field are more likely to complete than students who are not pursuing any degree whereas student pursuing degrees in non-IT fields are less likely to do so. The exponentiated coefficients in the Cox model can be interpreted as the marginal effects on the hazard rate. So, for instance, a male student has an expected probability of completing the course that is 1.13 times more than for a female student. Similarly, a student who is not working is 1.48 times more likely to complete the course than a student who is working in a non-IT company.

Column 2 presents the results with only course interaction variables, all of which are statistically significant aside from the completion of the second activity within 24 hours of the first. As expected, having completed 10 activities means that a student is much more likely to complete the whole course - specifically, they are 79 times more likely to do so than a student who has not. The speed of completing the activities also mattered. Completing the first ten activities within a week meant that a student was 3 times more likely to complete than a student who did not. However, students who work through the course fast, completing at least five activities in three days and ten in a week, are more likely to finish the course. Somewhat surprisingly, the effect of completing just five activities at the time of the activity report being generated was significantly negatively associated with overall course completion. The final column combines both sets of covariates together but most of the variables, aside from gender, retain their statistical significance and estimated sign though there are changes in the estimated magnitudes.

The last row in Table 2 presents the concordance statistics for the three models providing us with the means of making comparisons of the goodness-of-fit of the three models. The statistics show that it is the course interaction variables that have the largest impact on improving the goodness of fit, driving the concordance statistic up from 0.57 to 0.84. Including both demographic and course interaction variables has only a marginal improvement, going up to 0.86.

5. Discussion and Conclusion

This paper uses data from a MOOC on the open.uom.lk platform to study course completion and the time taken to do so using survival analysis methods, which has the advantage over standard classification models of using information on the timing of course completion and accounting for censored data to provide better estimates of the effect of covariates on the probability of course completion. Specifically, using the activity logs and registration survey for the introductory web design course in the Trainee Full Stack Developer program, we exploit both details of course interactions as well as demographic characteristics to predict the probability of completing the course within a given time frame. The results highlight student characteristics that are significantly associated with course completion while also enabling a comparison of the explanatory power of demographic vs course interaction variables.

The results suggest that the course interaction variables are more important for predicting course completion than the demographic variables. Specifically, students who have completed more than 10 activities and completed the first five or ten activities in a relatively short period of time are more likely to complete the course sooner. Interestingly, completion of only five activities is strongly negatively associated with completion overall suggesting that there may be a bottleneck in the course between the 5th and 10th activities. Further analysis making use of additional details of activity completion could be carried out to ascertain this.

These findings are consistent with the literature. Work by Brooks et al. (2015) find course interactions are more predictive of course completion than demographic characteristics in MOOCs, though demographic factors have been used to predict outcomes in traditional higher education systems. There is also evidence that students who maintain momentum during an online course are more likely to complete the course (Jiang et al. 2014; Bitkimirov and Klassen, 2008; Baugher, Varanelli and Weisbord, 2003). This implies that those at risk of dropping out could be flagged within the system and perhaps targeted with different types of interventions. However, the interventions must be carefully thought out - work by Borrella et al. (2021) suggests that while ad-hoc email reminders are not effective, interventions that make course content more approachable have a positive impact on course completion.

Among the demographic characteristics, gender is only marginally significant but educational background and employment status are significantly associated with completion. For example, students with a degree in IT are more likely to complete than students who are not following any degree. This is intuitive - these students are already motivated and have the foundation to learn the technical skills required for web development more quickly. On the other hand, students who are not employed are quicker to finish the course than those who are, particularly more so than those working outside the IT domain or are self-employed. This could be due to fewer time commitments on the part of those who are not employed. Perhaps surprisingly, students who are yet to do the Ordinary Level examination are more likely to finish the course than those with higher levels of secondary education. A potential explanation is that this group has a greater interest in the subject matter, where among the higher levels of education, there is a wider range of interest and experience in IT. The demographic characteristics identified as being associated with course completion suggest that the course is already positioned as one that is useful to those seeking to build their skills for entering into the IT industry.

The analysis presented here is subject to limitations. For instance, the class imbalance between completing and non-completing students is only partially addressed by selecting students who have completed at least the first activity. Using additional techniques to address the imbalance could improve the model further. The analysis is also constrained in terms of the available background characteristics of the participants, including their programming skills, though this may be partially captured by the variables indicating IT as the domain of higher education and employment. The analysis uses only a limited number of course interaction variables. Given the finding of their importance in predicting course completion, further work to identify bottlenecks in the course and behavioural patterns linked to faster completion of this course as well as the overall TFSD program should be carried out.

References

Alhothali, A., Albsisi, M., Assalahi, H., and Aldosemani, T. (2022) “Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review. Sustainability”, *Sustainability* 2022,14, 6199.

Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J. and Radi, N. (2017) “Machine learning approaches to predict learning outcomes in Massive open online courses”, International Joint Conference on Neural Networks (IJCNN).

Baugher Varanelli Weisbord, D.A.E. (2003) “Student hits in an internet-supported course: How can instructors use them and what do they mean?”, *Decision Sciences Journal of Innovative Education*, 1(2): pp 159-179.

Biktimirov, E.N. and Klassen, K.J. (2008) “Relationship between use of online support materials and student performance in an introductory finance course”, *Journal of Education for Business*, 83(3): pp 153-158.

Borrella, I., Caballero-Caballero, S. and Ponce-Cueto, E. (2022) “Taking action to reduce dropout in MOOCs: Tested interventions”, *Computers & Education*, 179(104412).

Brooks, C., Thompson, C. and Teasley, S. (2015) “Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses (MOOCs)”, 2nd ACM Conference on e-Learning.

Chaplot, D. S., Rhim, E., and Kim, J. (2015) “Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks”, *AIED 2015 Workshop Proceedings* - Vol 3

Damianov, D.S., Kupczynski, L., Calafiore, P., Damianova, E.P., Soydemir, G. and Gonzalez, E.(2009) “ Time spent online and student performance in online business courses: A multinomial logit analysis”, *Journal of Economics and Finance Education,* 8(2): pp 11-22.

Gitinabard, N., Khoshnevisan, F., Lynch, C.F. and Wang, E.Y. (2018) “Your actions or our associates? Predicting certification and dropout in MOOCs with behavioral and social features”, International Educational Data Mining Society.

Goel, M.K., Khanna, P. and Kishore, J. (2010) “Understanding survival analysis: Kaplan-Meier estimate”, *International Journal of Ayurveda Research*, 1(4): pp 274.

Greene, J.A., Oswald, C.A. and Pomerantz, J. (2015) “Predictors of retention and achievement in a massive open online course” , *American Educational Research Journal*, 52(5): pp 925-955.

Impey, C. and Formanek, M. (2021) “MOOCS and 100 Days of COVID: Enrollment surges in massive open online astronomy classes during the coronavirus pandemic”, *Social Sciences & Humanities Open*, 4(1).

Information and Communication Technology Agency (ICTA) of Sri Lanka. (2019) National IT-BPM Workforce Survey, 2019, TEAMS (Pvt) Limited.

Jiang, S., Williams, A., Schenke, K., Warschauer, M. and O'Dowd, D. (2014) “Predicting MOOC performance with week 1 behavior”, International Conference on Educational Data Mining.

Labrador, M.M., Vargas, G.R.G., Alvarado, J. and Caicedo, M. (2019) “Survival and risk analysis in MOOCs” *Turkish Online Journal of Distance Education*, 20(4): pp 149-159.

Matz, S.C., Bukow, C.S., Peters, H. (2023) “Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics”. *Sci Rep* 13, 5705.

Moreno-Marcos, P.M., Alario-Hoyos, C., Muñoz-Merino, P.J. and Kloos, C.D. (2019) “Prediction in MOOCs: A review and future research directions”, *IEEE transactions on Learning Technologies*, 12(3): pp 384-401.

Rulinawaty, R., Priyanto, A., Kuncoro, S., Rahmawaty, D. and Wijaya, A. (2023) “Massive Open Online Courses (MOOCs) as Catalysts of Change in Education During Unprecedented Times: A Narrative Review”, *Journal Penelitian Pendidikan IPA*, 9: pp 53-63.

Therneau, T.M. and Watson, D.A. (2017) “The concordance statistic and the Cox model”, *Department of Health Science Research*, 85: pp 1-18.

Ulfa, S. and Fatawi, I. (2012) “Predicting factors that influence students’ learning outcomes using learning analytics in an online learning environment”, *International Journal of Emerging Technologies in Learning*, 16(1): pp 4-17.

Van Der Net, J.B., Janssens, A.C.J., Eijkemans, M.J., Kastelein, J.J., Sijbrands, E.J. and Steyerberg, E.W. (2008) “Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies”, *European Journal of Human Genetics*, 16(9): pp 111-1116.

Vengroff, R. and Bourbeau, J. (2006) “In-class vs. on-line and hybrid class participation and outcomes: Teaching the introduction to comparative politics class”, APSA Teaching and Learning Conference.

Villano, R., Harrison, S., Lynch, G. and Chen, G. (2018) “Linking early alert systems and student retention: a survival analysis approach”, *Higher Education*, 76: pp 903-920.

Wintermute, E.H., Cisel, M. and Lindner, A.B. (2012) “A survival model for course-course interactions in a Massive Open Online Course platform”, *PloS one*, 16(1).