Submissions Abstract Book - All Papers (All Submissions)

0254

P1 | Beaumaris Lounge
Chaired by Jacqueline Stevenson

Thu 12 Dec 2019

16:45 - 17:15

Predicting Progression to Postgraduate Study in UK Higher Education using Machine Learning Techniques

Carolina Cuesta-Lazaro[1], Pep Mateos Gonzalez[2], Arnau Quera-Bofarull[1], Paul Wakeling[2]

[1]Durham University, Durham, United Kingdom [2]University of York, York, United Kingdom

Research Domain: Access and widening participation (AWP)

Abstract:

We develop an accurate model that predicts progression to postgraduate study in UK higher education using machine learning, assessing its potential in the field of educational inequalities. To answer our research question – "*which characteristics of graduates predict progression to postgraduate education?"*– we analyse a large dataset (N=1,361,003) containing information on all first-degree graduates in UK higher education from the academic years 2012/13 to 2016/17 and their post-study destinations. Applying machine learning permits testing of the relevance of certain students' characteristics which previous research has identified to have an effect on progression to postgraduate study and to 'reveal novel relationships among the variables' (Delen, 2010, p. 499). Research suggests that subject, gender, ethnicity, social class and type of institution attended affects progression (Wakeling and Hampden-Thompson, 2013). Descriptive analysis of our dataset supports this. While these results are promising, formal modelling is needed to decipher which characteristics better predict progression.

Paper:

Context and purpose of this paper

This paper aims to develop an accurate model that predicts progression to postgraduate study in UK higher education using machine learning techniques, assessing the potential of the latter in the field of higher education research and its implications for policy-making. Our core research question, that is "*what are the characteristics of graduates that are more likely to progress to postgraduate education"*, is a relevant one in terms of current debates on inequalities in access to higher education. Recent expansion of postgraduate enrolments has coincided with growing concerns that improvements in access to undergraduate higher education will be eclipsed by rising inequalities in access to postgraduate education. As postgraduate study and the acquisition of higher graduate qualifications is becoming increasingly important in shaping individuals' wages (Lindley and Machin,

2013; Wakeling and Laurison, 2017), postgraduate education may effectively become the 'new frontier of social mobility' (Lampl, 2013). Notwithstanding, while there has been a recent proliferation of research exploring inequalities in access to postgraduate education (cf. Wakeling, 2005; Stuart *et al.*, 2008; Morgan, 2014), there is still scope to better understand the factors that influence progression to postgraduate study.

Data and methods

This paper does so by using machine learning techniques to analyse a large and rich dataset (N=1,361,003) commissioned to the UK's Higher Education Statistics Agency (HESA) that contains information of all undergraduate leavers in UK higher education that graduated between the years 2012/13 to 2016/17 and their post-study destinations, including postgraduate study. Graduates' information found in this dataset includes data on their social backgrounds –such as parental education, ethnicity or their socio-economic classification (Rose, Pevalin and O'Reilly, 2005)–, their attainment prior to and during higher education, and the institutions and subjects of their undergraduate education and postgraduate education. The application of machine learning to this dataset will allow us to predict progression to postgraduate study using techniques that 'outperform standard regression models' (Hindman, 2015, p. 48).

Machine learning is a novel application of artificial intelligence to the analysis of rich data that uses learning algorithms to identify patterns even in the presence of complicated nonlinear interactions (Bzdok, Altman and Krzywinski, 2018). Machine learning has been proven to be a powerful predicting technique compared to frequentist statistics in a variety of research areas, including medical image processing (Bullock, Cuesta-Lazaro and Quera-Bofarull, 2018), political orientation (Colleoni, Rozza and Arvidsson, 2014) and student attrition (Delen, 2010). This is particularly the case with datasets with large numbers of explanatory variables. Machine learning allows to determine which variables are more important in predicting an outcome – in this case whether a graduate progresses to postgraduate study– and does so randomly dividing the data into sets, developing prediction models with one of the sets and then computing prediction errors using a cross-validation and a test set (Hindman, 2015). It also uses a "shrink approach", that is removing explanatory variables that only provide small increases in accuracy, which makes the model more manageable and reduces overfitting (ibid.).

Preliminary findings and research plan

Research designs that include machine learning permit to test the relevance of certain students' characteristics that previous research has identified to have an effect on progression to postgraduate study and to 'reveal novel relationships among the variables' (Delen, 2010, p. 499). Previous research in the UK suggests that gender, ethnicity, social class and type of institution attended affects progression to postgraduate study (Wakeling, 2005; Wakeling and Hampden-Thompson, 2013). Descriptive analysis of our dataset supports this. Notwithstanding, this analysis also indicates that the relationship between the latter characteristics and progression to postgraduate study weakens, particularly after the academic year 2016/17, when student loans to support postgraduate education were introduced (Hubble, Foster and Bolton, 2018). Moreover, our research reveals clear variance in progression rates among students with different attainment levels, both prior to higher education and in their undergraduate education.

While these results are promising, both sociologically and for policy-making, formal modelling is needed to decipher which students' characteristics better predict progression to postgraduate education. We plan to do so by applying machine learning techniques, which in turn will allow us to explore their potential for higher education research.

References

Bullock, J., Cuesta-Lazaro, C. and Quera-Bofarull, A. (2018) 'XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets', *CoRR*, abs/1812.0. Available at: http://arxiv.org/abs/1812.00548.

Bzdok, D., Altman, N. and Krzywinski, M. (2018) 'Statistics versus machine learning', *Nature Methods*, 15(4), pp. 233–234.

Colleoni, E., Rozza, A. and Arvidsson, A. (2014) 'Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data', *Journal of Communication*, 64(2), pp. 317–332.

Delen, D. (2010) 'A comparative analysis of machine learning techniques for student retention management', *Decision Support Systems*. Elsevier B.V., 49(4), pp. 498–506.

Hindman, M. (2015) 'Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences', *Annals of the American Academy of Political and Social Science*, 659(1), pp. 48–62.

Hubble, S., Foster, D. and Bolton, P. (2018) *Postgraduate loans in England*. 7049. London: House of Commons Library.

Lampl, P. (2013) *We need the best postgrads, not just the richest*. Available at: https://www.suttontrust.com/newsarchive/need-best-postgrads-just-richest/ (Accessed: 27 February 2019).

Lindley, J. and Machin, S. (2013) *The Postgraduate Premium. Revisiting Trend in Social Mobility and Educational Inequalities in Britain and America*. London: The Sutton Trust.

Morgan, M. (2014) 'Patterns, drivers and challenges pertaining to postgraduate taught study: an international comparative analysis', *Higher Education Research and Development*, 33(6), pp. 1150–1165.

Rose, D., Pevalin, D. J. and O'Reilly, K. (2005) *The National Statistics Socio-economic Classification: Origins, Development and Use*. Basingstoke: Palgrave Macmillan.

Stuart, M. *et al.* (2008) *Widening participation to postgraduate study: Decisions , deterrents and creating success. Project report*. York: The Higher Education Academy.

Wakeling, P. (2005) 'La noblesse d'état anglaise? Social class and progression to postgraduate study', *British Journal of Sociology of Education*, pp. 505–522.

Wakeling, P. and Hampden-Thompson, G. (2013) *Transition to higher degrees across the UK: an*

analysis of national, institutional and individual differences. York: The Higher Education Academy.

Wakeling, P. and Laurison, D. (2017) 'Are postgraduate qualifications the "new frontier of social mobility"?', *The British Journal of Sociology*, 68(3), pp. 533–555.