Using K nearest neighbours to evaluate the fairness of LSE's degree classification rules (for MSc Economics).

Matthew Bradbury, Miles McCollum

See this project presented by the researchers: https://youtu.be/pf7tnDlfHVM
If this project has informed your practice, let us know at lse.changemakers@lse.ac.uk

The set of rules used to classify degrees awarded at LSE is different to the vast majority of universities in the UK. Put simply, a master's degree is split into four equally-weighted grades, and in order to earn a first class classification a student must either attain a score of 70 in three of these, or do so in two whilst averaging at least 68 in the remaining grades. When this system is compared to the one used in most other universities, wherein the benchmark is simply to average 70 across the entire summative part of the degree, one must ponder the benefits of being different.

The gist of our research was to analyse whether the current degree classification system awards the same degrees to the same 'type' of student each year in order to assure fairness. The data used in this study are the raw marks and degrees awarded to MSc Economics students in 2012, 2013 and 2018. In order to quantitatively classify students into a 'type', we used the machine learning algorithm of k-nearest neighbours (k-NN). In the context of our research, the method works as follows:

One year's exam results, consisting of four raw exam marks and a degree classification, is used as the 'training data'. Essentially, each student's set of marks is plotted as a point in four-dimensional space and the point itself is assigned the degree classification which that student was awarded that year.

Another year's raw marks are then plotted in the same four-dimensional space, without classification. K-NN then assigns each unclassified point a degree classification corresponding to the most popular degree amongst its k-nearest neighbours in the training data. Therefore the k-NN method groups students into 'types' depending on the geometric proximity of their marks. Take the following example for illustration:

K=11, and we are classifying a student from the 2013 cohort using the 2012 results. For this student, the 11 nearest 2012 students in four-dimensional space are isolated. Of these 11, six attained a merit (more than any other award) – therefore, the 2013 student in question will be awarded a merit also.

By extending this method to every student, using every academic cohort in turn as the training data, we gained a larger picture of how consistent the LSE rules are at awarding degrees, and whether students who achieve geometrically similar grades each year are given the same classification. An example of the output of the process is below:

LSE Classification Rules		Awarded grade 2013 cohort			
		Distinction	Merit	Pass	Fail
Machine-predicted grade using 2012 cohort	Distinction	9	1	0	0
	Merit	7	63	9	0
	Pass	0	14	36	18
	Fail	0	0	0	0

This table shows that, for example, of the 16 students who received a distinction, 9 had marks closer to the distinction candidates' from the previous year, while 7 had marks more closely resembling 2012 students who had been awarded a merit.

Evaluation

A degree classification system which is perfectly consistent would only yield entries in the diagonals of the above table – but it is unclear whether such a system exists, other than to use the k-NN process itself to award degrees (bad idea). The extent to which the LSE system is inconsistent is also unclear.

For evaluation, we run the process as above, but using the training data with degrees awarded using the more common university classification system instead of what LSE actually awarded (although there was a lot of overlap between these). The output from this using the same data is below:

Normal Classification Rules		Awarded grade 2013 cohort			
		Distinction	Merit	Pass	Fail
Machine-predicted grade using 2012 cohort	Distinction	10	0	0	0
	Merit	3	76	4	0
	Pass	0	0	46	17
	Fail	0	0	0	1

Comparing the two tables, we see that under the common classification rules there are fewer "inconsistent" (non-diagonal) entries. When this process was repeated using all 6 permutations of training and input data sets, we found that 22.1% of k-NN predictions were inconsistent under the LSE rules, compared to 14.4% under the common rules.

This measure of a classification rules' consistency - calculating proportion of non-diagonal entries - is crude. While beyond the scope of this study, there are test-statistics which can be yielded from the outputs and used for hypothesis testing, to see if the common classification system is significantly more consistent at awarding degrees than LSE's.

Expanding the data set to more years, more degrees and bachelors degrees also would improve the power of this research. These are considerations for future work on this topic.

Recommendation

Our research found that, judging by our metric of measurement, the rules which LSE employs to award degrees are less fair that those at other universities, in the sense that they are less consistent in the type of students to which they award certain classifications year-on-year. With this in mind, we recommend that the LSE commit further resources to the evaluation of its classification rules and increases the transparency concerning why rules are different at LSE.

We stop at this, rather than recommending to make the switch to the common classification rules for two reasons. Firstly, there are many other factors which determine the merit of a classification system beyond just its consistency, such as its popularity. Secondly, there are many other ways to measure its fairness, such as using another machine learning technique to k-NN. These two topics, along with many others, offer fruitful avenues for future study.