

# Project Proposal

## Explaining Grades with the use of Behavioural Data

In the run up to applying for university, teachers play a huge part in the application process by providing predicted grades for students. These are used as a rough guide by universities as to whether or not to provide an offer for students. However, these predictions are, in the majority, inaccurate, and in the case of disadvantaged students, tending to be lower than what is actually achieved.

In a report by the UCL Institute of Education, it was revealed that 'just one in six (16%) of predicted A-level results turn out to be correct' <sup>[1]</sup>, that is to say that the grade predicted is exactly the same as the one achieved. It also suggests that those grades that are under-predicted (one in ten) are more likely to come from those with disadvantaged backgrounds.

It has been suggested that teachers' predictions are heavily influenced by previous exam results, which can potentially introduce error into the predictions. However, it must be noted that humans are subjective; we don't have the tools to deal with such large amounts of information to make accurate calls without bias. So would it be possible to create a system that can help teachers identify anomalies and in some way suggest an appropriate prediction or grade range, based off factors other than just past exam results?

Mr. Chris Evans, the deputy head of Reading Grammar School, has been in contact and has a vast amount of data from six different years that can be used. Included within this is scores for a variety of measures across all subjects to do with behavioural traits, such as curiosity and perseverance. In addition, there is also the potential to have access to attendance for each student, and their exam results for these years.

This data is clearly very personal, and any sort of data sharing without proper regulation would cause a huge issue. Following a discussion with Mr. Evans, he is also aware of this and will be in close contact to ensure all regulations and protocols to do with the ethics of sharing personalised data are followed.

The main aims would be to statistically explore the data. Data mining will play a huge part in this, exploring a range of machine learning techniques to find patterns in the data. Possible exploration areas include, but are not limited to <sup>[2]</sup>:

- **Linear models** – These work well when involved with numeric data, as is the case for this model specific data
- **Clustering methods** – These can be used effectively when looking to split the data into groups, which could be the case for this grade prediction
- **Probabilistic methods** - Upon analyzing the data, some features may display no linear relationship, but this can open the door to the plethora of probabilistic techniques available

Using these methods would ideally return a relationship between these aforementioned behavioural traits and the grades achieved by the students.

There has been a number of studies into so called 'Educational Data Mining' (EDM), which is based around making discoveries within the data from educational settings<sup>[3]</sup>. There are international conferences that bring together researchers to analyze large data sets, and the proceedings from the 2008 conference contains a number of papers that display the various investigations that have been undertaken. <sup>[4]</sup>

While he is providing the data, Mr. Evans is also extremely interested in the result of this study. He is under the impression that such a system can only benefit schooling, and is confident that it could be implemented within his school and others.

[1] Call for university application overhaul (2016).

<https://www.ucl.ac.uk/ioe/news-events/news-pub/dec-2016/report-reveals-just-16-percent-of-predicted-A-level-results-are-correct>

[Accessed 11th May 2017]

[2] Witten, I., Frank, E., Hall, M., Pal, C. " Data Mining: Practical Machine Learning Tools and Techniques".

[https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+algorithms+data+mining&ots=8HIMsgnCy8&sig=0IRitlvtXZ-](https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+algorithms+data+mining&ots=8HIMsgnCy8&sig=0IRitlvtXZ-Wx7uSkRyff1y4W6o#v=onepage&q=probabilistic%20methods&f=false)

[Wx7uSkRyff1y4W6o#v=onepage&q=probabilistic%20methods&f=false](https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=machine+learning+algorithms+data+mining&ots=8HIMsgnCy8&sig=0IRitlvtXZ-Wx7uSkRyff1y4W6o#v=onepage&q=probabilistic%20methods&f=false)

[Accessed 11th May 2017]

[3] Data Mining for Education.

<http://users.wpi.edu/~rsbaker/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf>

[Accessed 11<sup>th</sup> May 2017]

[4] The 1<sup>st</sup> International Conference on Educational Data Mining Montreal, Quebec, Canada, June 20-21, 2008 Proceedings. (2008).

<http://www.educationaldatamining.org/EDM2008/uploads/proc/full%20proceedings.pdf>

[Accessed 11<sup>th</sup> May 2017]