IJMRE paper

Florence Gabriel, Jason Signolet, Martin Westwell

Department, University, City, Country

Provide full correspondence details here including e-mail for the corresponding author

Provide short biographical notes on all contributors here if the journal requires them.

Repeat the title of your article here

Type or paste your abstract here as prescribed by the journal’s instructions for authors. Type or paste your abstract here as prescribed by the journal’s instructions for authors. Type or paste your abstract here as prescribed by the journal’s instructions for authors. Type or paste your abstract here.

Keywords: word; another word; lower case except names

Subject classification codes: include these here if the journal requires them

# Introduction

* Interdisciplinary problems:
* May bring together concepts and variables that may never have been considered part of the same framework before.
* Variables may have been considered before, but not in a holistic manner
* Possible lack of theory combining them
* Difficult to generate hypotheses without adequate theory
* Machine learning tools can be used for data-driven hypothesis generation
* Crunch large numbers of variables and objects simultaneously
* Automatically identify interactions in the data
* With the right know-how, as easy to implement as classical stats tools

In the present study, we give a demonstration of how modern machine learning techniques can allow researchers to extract useful information from large data sets such as these by analysing multiple dimensions at once and modelling non-linear relationships. Here, we apply these techniques to the PISA 2012 assessment, which focussed on mathematical literacy.

## Intro to machine learning

* What is machine learning and how does it differ from statistical and mechanistic models
* Types of ML
  + Supervised
    - Regression
    - Classification
  + Unsupervised
    - Clustering
    - Anomaly detection
* How does it work
  + Attempt to emulate real-life results as closely as possible
  + Data-driven
    - Good models require good data
  + Computer uses optimisation algorithms to find the best predictive combinations of features and parameters for a model
    - Iterative process
    - Optimisation algorithms use search techniques to find the combinations of features and parameters that produce the lowest error
* Who uses it
  + Data scientists, comp sci
  + Collaborations with researchers and educators
  + Useful to learn, just like learning how to do stats
* Why is it useful here
  + Highly intuitive
  + Good for large datasets
  + Can give you data-driven hypotheses
    - Can give predictions and recommendations
    - Especially useful at interdisciplinary boundaries where there are large volumes of data without a mature theoretical framework.
* Easy to determine if the model is any good by using domain knowledge, Cross validation, bootstrapping, test sets.

## Using ML to model the relationship between dispositions, demographics and mathematical literacy

* Focus on PISA 2012 as an example
  + Not big data, but a decent quantity and quality
* Address questions at the discipline boundary between psychology and education
  + Disposition, demographics and performance

Insist on interdisciplinarity

* Introduce RF and GBM
* Xgboost is best-in-class for this type of problem (see Kaggle)
  + Demonstrate predictive power in non-linear systems
    - Advantages over linear regression
  + Demonstrate identification of important variables
    - Relative influence
    - Partial dependence plots
  + Demonstrate interaction discovery
    - Can we identify known interactions?
    - Can we find new interactions?

We analysed the Australian data from the PISA 2012 assessment to investigate the relationship between disposition and achievement in mathematics in Australian high-school students. PISA is an international student survey conducted every three years by the Organisation for Economic Co-operation and Development (OECD). The data set is publicly available, has good coverage of all Australian jurisdictions and contains data on the demographics, mathematics dispositions and mathematics performance of over 14,000 fifteen-year-old students. In the following section we will describe the dispositions towards mathematics used in the PISA 2012 survey. These included: mathematics anxiety; mathematics self-concept; mathematics self-efficacy; perceived control in school; perceived control in mathematics; extrinsic motivation; intrinsic motivation; and subjective norms. Demographics included gender, indigenous status, socio-economic status (SES), and state.

### Dispositions

There is a wealth of evidence indicating that self-efficacy and self-concept are the strongest correlates of academic achievement and participation 3-5. Self-efficacy and self-concept are highly related yet distinct constructs, both related to self-evaluation 6. Self-efficacy is the belief one has that they are able to complete a specific task or succeed in a specific situation 7 (e.g., “How confident do you feel about calculating how many square metres of tiles you need to cover a floor?”), whereas self-concept is an individual’s beliefs about their achievements, abilities or skills in a particular area of competence 8 (e.g. “Do you agree with the statement, *I learn mathematics quickly*?”).

Mathematics anxiety is a disposition where there is a clear, well-reported demographic-driven difference, namely girls are generally more strongly affected than boys 9. Mathematics anxiety can be defined as “feelings of fear, apprehension, or dread that many people experience when they are in situations that require solving math problems” 10. It has a negative influence in participation and lifelong learning in mathematics 11.

Perceived control is the sense that an individual has of being able to influence the events/situations that they face. PISA 2012 surveyed students’ perceived control in mathematics classes and in school in general. It has been shown to be a negative predictor of anxiety 12,13 and a positive predictor of academic success 14.

The PISA 2012 assessment surveyed two types of motivation: intrinsic and extrinsic. Intrinsic motivation (called mathematics interest in PISA 2012) is the drive that an individual has to do an activity, simply because they enjoy, or are interested in, doing said activity 15. In contrast, extrinsic motivation (called instrumental motivation in PISA 2012) is the drive to do an activity in order to achieve some goal (e.g. a student may be motivated to pursue mathematics if they believe that mathematics is of benefit to their future studies or employment) 16. In mathematics education, intrinsic motivation has been shown to be positively related to achievement, whereas the relationship between extrinsic motivation and achievement is much less clear-cut, with some studies showing weak positive correlations, while others have even indicated a negative relationship 17,18.

In the PISA 2012 assessment, subjective norms relate to the importance that the individual’s parents and friends place on studying/using/doing mathematics 1. This disposition is thought to affect both intentions and behaviour 19.

### Demographics

On average, students with indigenous status in Australia are among the most disadvantaged in terms of educational outcomes, both compared to non-indigenous Australians and to other indigenous populations around the world 20,21. Compared to non-indigenous students, Australian indigenous students are more likely to be in the lowest SES quartile 21. These students report low academic self-concept 20.

A recent meta-analysis of 242 studies published between 1990 and 2007 found little evidence for an overall difference in mathematics performance between male and female students 22. Another meta-analysis focussed on TIMSS 2003 and PISA 2003 comparing international data sets also reported no difference between the genders in mathematics performance as measured by TIMSS, but they found a slight difference in favour of boys in mathematical literacy as measured by PISA 23.

SES has been positively correlated with numeracy, both at the level of the student and the level of the school 24. According to the Australian Council for Educational Research (ACER) PISA 2012 report on Australia, the difference in mathematical literacy between students in the highest and lowest SES quartile was, on average, equivalent to around two and a half years’ worth of schooling 25.

Australia’s federal system includes six states and two territories (hereafter simply referred to as states), to which the primary responsibility of school education is devolved 26. The state and federal governments share priorities and agree initiatives on a national level via consultative arrangements such as the Council of Australian Governments (COAG) and the Ministerial Council on Education, Early Childhood Development and Youth Affairs (Education Council) 26. ACER reported significant differences in mathematical literacy between the states, with the Australian Capital Territory and Western Australia outperforming the other states, and the Northern Territory being outperformed by all other states 25.

# Heading 1: use this style for level one headings

Paragraph: use this for the first paragraph in a section, or to continue after an extract.

New paragraph: use this style when you need to begin a new paragraph.

Display quotations of over 40 words, or as needed.

* For bulleted lists

1. For numbered lists

Displayed equation ( )

## Heading 2: use this style for level two headings

### Heading 3: use this style for level three headings

*Heading 4: create the heading in italics*. Run the text on after a punctuation mark.

Acknowledgements, avoiding identifying any of the authors prior to peer review

1. This is a note. The style name is Footnotes, but it can also be applied to endnotes.

References: see the journal’s instructions for authors for details on style

Table 1. Type your title here.

Figure 1. Type your title here. Obtain permission and include the acknowledgement required by the copyright holder if a figure is being reproduced from another source.