Learning from Machines to Close the Gap between Funding and Expenditure in the Australian National Disability Insurance Scheme

Satish Chand, Yu Zhang

School of Business, University of New South Wales at Canberra

Abstract

The Australian National Disability Scheme (NDIS) allocates funds to participants for purchase of services contained in their officially approved plans.

Just one percent of the sample of 70,000 NDIS participants were able to spend more than 90 percent of their allocated budget with two percent (i.e., 1,405 participants) having failed to spend any, meaning that most of the participants were left with unspent funds. The gap between the allocated budget and realised expenditure can be due to errors in both, which we attempt to close using machine learning techniques.

Four machine learning models were evaluated with three experiments designed to test the efficacy of each model, compare the rates of learning between humans and machines, and investigate the explanatory variables chosen by the models. The results show that: (i) machines learn ``faster'' than humans; (ii) selection of an appropriate machine learning model can improve the accuracy and efficiency of the NDIS funding allocation; and (iii) the significant explanatory variables used across the models are similar.

Emerging challenge for the NDIS

The NDIS has made significant progress in improving access to 'reasonable and necessary supports for people with disability'. The Scheme is projected to provide funding of $22b to approximately 500,000 Australians every year within the next five years. This level of commitment to a public insurance scheme that surpasses Medicare is unprecedented in Australia [1]. An emerging challenge for the NDIA is to allocate funding to need but within the resource envelope. On this, the NDIS Costs Study Report observes that:

(i) "…too much focus has been on quantity (meeting participant intake estimates) and not enough on quality (planning processes)…";

thus,

(ii) "…the NDIA must find a better balance between participant intake, the quality of plans, participant outcomes, and financial sustainability…".

Data and methods

The objectives of this research are to explain the heterogeneity in utilisation rates using available quantitative techniques from artificial intelligence.

Data: Overall 6 quarters of NDIS data is used including two quarters in 2019 (Q3 and Q4 of 2019) and all four quarters in 2020. There are 38,155 observations with 10 attributes for each record in this dataset after removing missing values and duplicate characteristics [2].

Experiments:

1. Four Machine Learning techniques, i.e., Linear Regression, Support Vector Machines (SVM), Decision Tree and Multilayer Perceptron (a class of feed-forward neural network, short for MLP) are employed to estimate the allocation of funding (Budget) and utilisation of funding (Expenditure). The efficacy of the techniques are evaluated in terms of their predicted to actual outcome – see figures right.
2. Quarterly Budget prediction for Q1, Q2, Q3 and Q4 of 2020 is made separately using actual historical data and rolling forecasts, respectively. This experiment is designed to assess the ability of machines to learn to close the Gap between the Budget and Expenditure over the quarters compared to that of human policymakers; this is quantified as Rate of Learning and elaborated on later.

Main finding 1

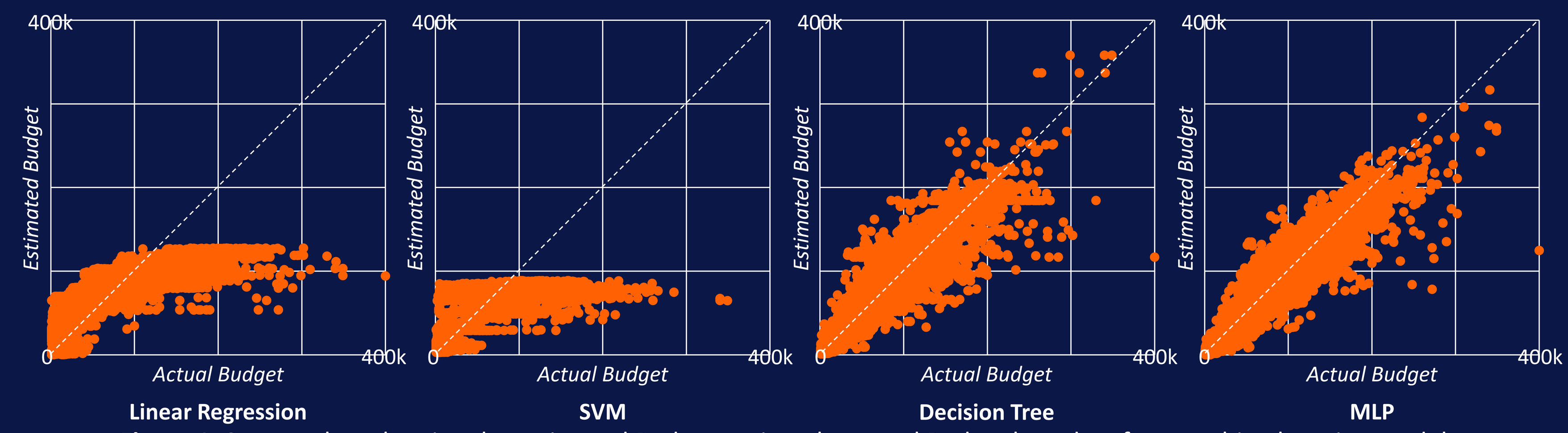


Figure 1. Scatter plots showing the estimated Budget against the actual Budget based on four machine learning models.

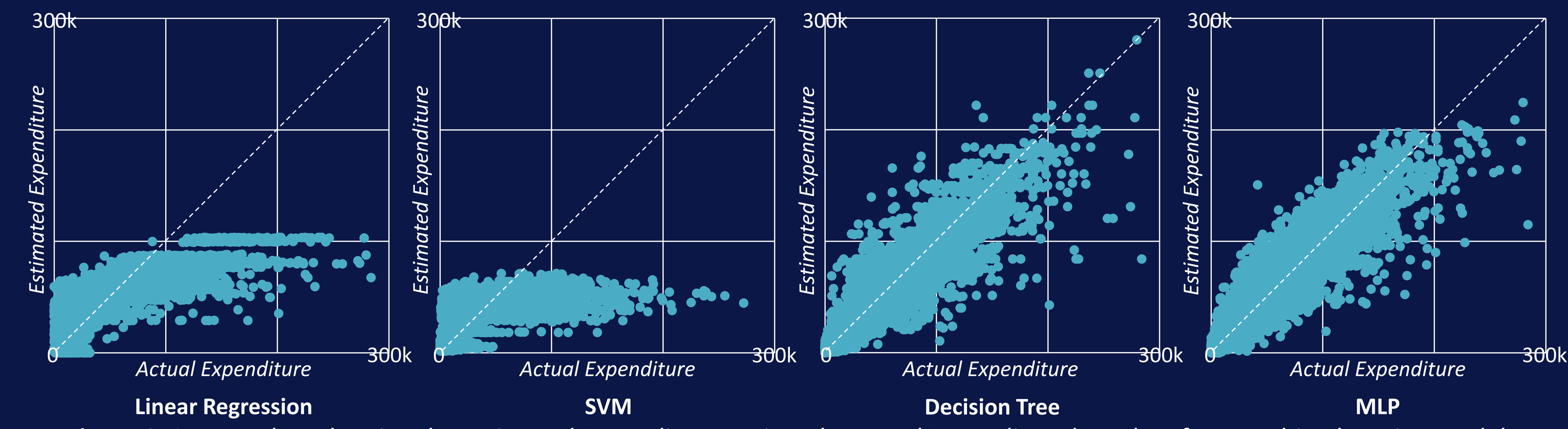


Figure 2. Scatter plots showing the estimated Expenditure against the actual Expenditure based on four machine learning models.

- The figures show that the Budget and Expenditure estimated by the Decision Tree and Neural Network models outrank the other two models in terms of performance.
- Compared between the results of MLP and Decision Tree, the Decision Tree has achieved the results that are more scattered than that from the MLP model. This is possibly due to the fact that the MLP model is better at handling the extreme cases in terms of estimating these variables.
- These figures also indicate how machines can effectively estimate the allocation of funding and expenditure of the participants based on the historical data, which can automate some of these functions sparing humans to spend their efforts at reviewing and approving applications.

Main finding 2

A decision tree model is trained to predict the Budget for Q1, Q2, Q3 and Q4 of 2020 using NDIS data [2]. Specifically, the data of Q3 and Q4 of 2019 was initially used to train the model for predicting the Budget in Q1 of 2020. For 2020 Q2 prediction, the training dataset is expanded by adding the data of Q1 of 2020. The same rule applies to the following predictions for Q3 and Q4 of 2020. Two methods of expending training datasets are employed:

- ***Updated Prediction*** - includes the actual historical data of one more quarter. For instance, the actual data of 2020 Q1 is added to the training dataset together with 2019 Q3 and Q4 data to predict the Budget of 2020 Q2.
- ***Autonomous Prediction*** - uses the predicted values for Q1 2020 instead of the actual value for the quarter. The use of predicted values for the model instead of the actual values for the rolling forecast is used to assess rates of learning of the machine compared to policymakers.

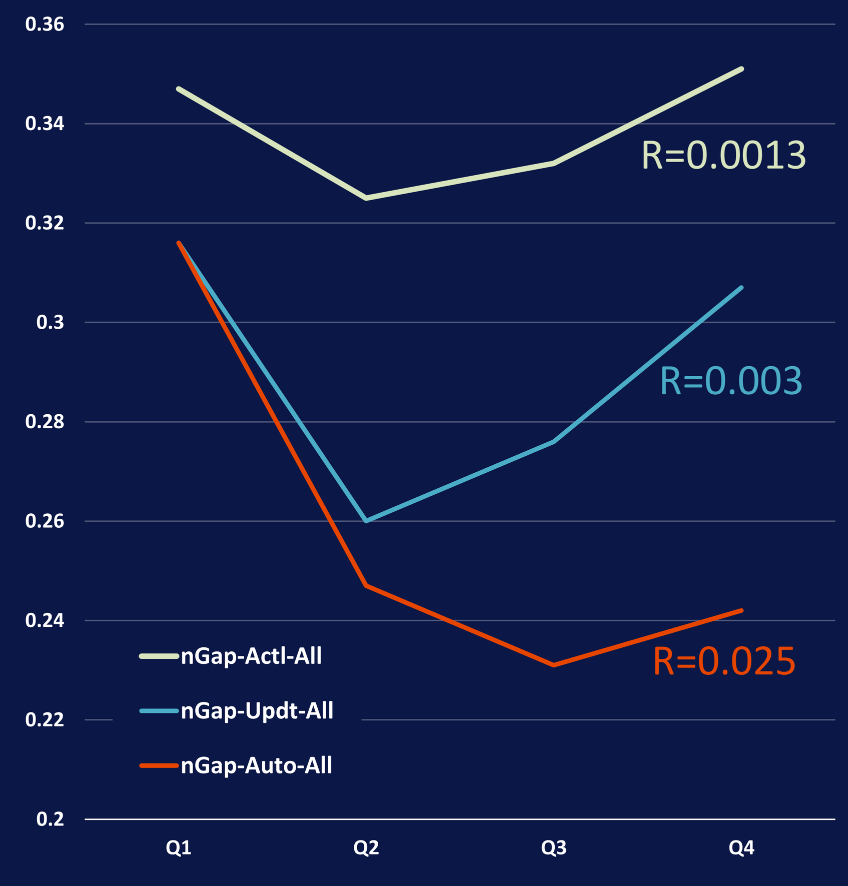


Figure 3. Comparison of the normalised Gaps that are generated based on actual values, Updated Prediction and Autonomous Prediction.

Figure 3 shows that both predicting methods outrank humans in closing the gap, and this gap is closed faster when machines roll their forecasts based on the past forecasts; that is, are left to forecast autonomously!

The Rate of Learning, R, is defined as:

$$R = \left|\frac{\Delta nGap}{\Delta Q}\right|$$

where $\Delta nGap$ refers to the change of nGap over the quarter interval, ΔQ. The larger this R is, the more effective its associated learning method has been. With continuous learning, the $nGap$ is expected to drop and gradually reach to a stable level, meanwhile ΔR will reach 0.

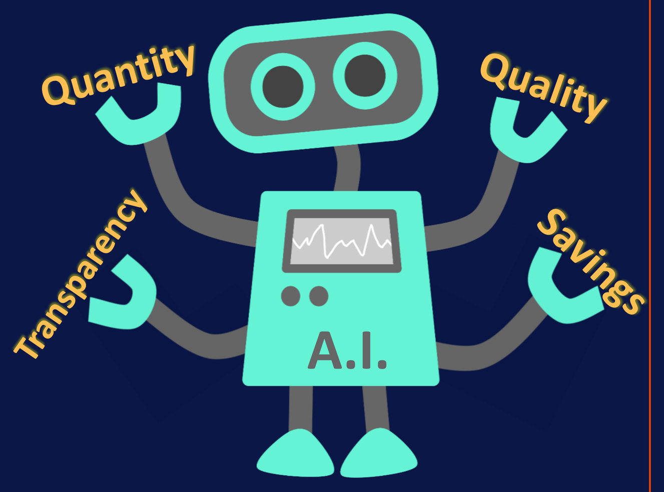
The R for human planners and two predicting methods is listed at the tails of the curves, indicating that hardly any learning took place over time by the planners, however the machines were able to pull down the $nGap$ relative to the actual budgets, and the Autonomous Prediction did the best.

Conclusion and Contribution

This research demonstrates that machine learning techniques can be employed to automate the estimation of the budget and expenditure of the participants in the NDIS using existing information, releasing humans to focus on fine-tuning the plans.

Moreover, machines learn faster than humans when it comes to closing the gap between the budgets and expenditure over time, helping improve the quality of NDIS plans and reducing under-spending by the participants.

In summary, the NDIS policymakers can benefit by automating some of the manual processes, leaving room for more personalised interventions. The fact that machines learn faster than humans, and can do so transparently at lower costs, has the potential to improve resource allocation.



Contact

Prof. Satish Chand
UNSW at Canberra
Email: s.chand@adfa.edu.au

Dr. Yu Zhang
UNSW at Canberra
Email: yu.zhang@adfa.edu.au

References

[1] Australian Productivity Commission (2017) Productivity commission study report: National disability insurance scheme (NDIS) costs.
[2] NDIS (2021) NDIS data downloads. URL: https://data.ndis.gov.au/data-downloads.