Better member outcomes in superannuation through data driven financial literacy prediction

Ben Culbert

Submission Date: 8 Jun 2021

Supervisor: Prof. Guandong Xu
guandong.xu@uts.edu.au

Co-supervisor: Dr. Shaowu Liu
shaowu.liu@uts.edu.au

University of Technology Sydney

Faculty of Engineering & IT, School of Computer Science

Advanced Analytics Institute

Certificate of original authorship

I, Ben Culbert declare that this thesis, submitted in fulfilment of the requirements for the award of Doctorate in Analytics, in the School of Computer Science, Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This thesis has not been submitted for qualifications at any other academic institution. This research is supported by an Australian Government Research Training Program Scholarship.

Signature:

Production Note:
Signature removed prior to publication.

Date:

30 May 2021

Acknowledgement

I wish to express my great appreciation to those who have supported me throughout my candidature.

I thank my supervisor Professor Guandong Xu for his constant support and unwavering belief in me and the value of the work we do together.

I wish to thank my colleagues at Colonial First State. In particular, I offer thanks to Todd Stevenson and James Brownlow for creating this opportunity for me and offering me the time and freedom to explore the greatest depths of this discipline and gain new skills. Thank you also to Charles Chu for his grounded advice and calming influence.

Finally, I wish to express my deepest gratitude to my partner Tatiana, for her enduring support, motivation and optimism, and for the sacrifices she's made in enabling me to complete this endeavour.

List of Publications

Culbert, B., Fu, B., Brownlow, J., Chu, C., Meng, Q. & Xu, G. 2018, 'Customer Churn Prediction in Superannuation: A Sequential Pattern Mining Approach', *Australasian Database Conference*, Springer, pp. 123-34.

Culbert, B., Brownlow, J. & Xu, G. 2021, 'Predicting financial literacy using debiased multi-output regression', *Journal of Behavioural and Experimental Finance*, [In Review]

Culbert, B., Liu, S. & Xu, G. 2021, ' Financial literacy in superannuation: Insights and evidence for positive financial outcomes and decision making ', Journal of Pension Economics and Finance, [In Review]

Chu, C., Brownlow, J., Meng, Q., Fu, B., Culbert, B., Zhu, M., Xu, G. & He, X. 2017, 'Combining heterogeneous features for time series prediction', International Conference on Behavioural, Economic, Socio-cultural Computing (BESC), IEEE, pp. 1-2.

Brownlow, J., Chu, C., Xu, G., Culbert, B., Fu, B. & Meng, Q. 2018, „A Multiple Source Based Transfer Learning Framework for Marketing Campaigns", International Conference on Neural Networks (IJCNN)

Brownlow, J., Chu, C., Fu, B., Xu, G., Culbert, B. & Meng, Q. 2018, 'Cost-Sensitive Churn Prediction in Fund Management Services', International Conference on Database Systems for Advanced Applications, Springer, pp. 776-88.

Vo, N.N., Liu, S., Brownlow, J., Chu, C., Culbert, B. & Xu, G. 2018, 'Client Churn Prediction with Call Log Analysis', International Conference on Database Systems for Advanced Applications, Springer, pp. 752-63.

Bateman, H., Brownlow, J., Culbert, B., Chu, C., Eckert, C., Fu, B. & Thorp, S. 2019, ARC Centre of Excellence in Population Ageing Research Industry Report 2019/2.

List of tables and figures

Table 2.1	Financial literacy questions - 2004 HRS	16
Table 2.2	Confusion matrix	27
Figure 3.1	Survey response rates	37
Figure 3.2	Basic financial knowledge assessment results	37
Table 3.1	Basic financial knowledge questionnaire	38
Figure 3.3	Basic knowledge by demographic group	41
Figure 3.4	Superannuation knowledge histogram	42
Table 3.2	Superannuation knowledge questionnaire	42
Figure 3.5	Superannuation knowledge by demographic group	45
Figure 3.6	Composite financial knowledge assessment results	46
Table 3.3	Financial attitude	47
Table 3.4	Financial behaviour	48
Table 3.5	Self-assessment performance	51
Figure 3.7	Self-assessment error – Basic financial knowledge	51
Figure 3.8	Self-assessment error – Superannuation knowledge	52
Table 3.6	Correlation coefficients for subjective and objective knowledge	52
Table 3.7	Regression results – Risk, return and account balance	53
Table 3.8	Regression results, coefficient of determination	55
Table 3.9	Avg. literacy scores by contribution status (12 months)	56
Table 3.10	Classification performance – voluntary contributions	57
Table 3.11	Number of investment options per account	58
Table 4.1	Financial knowledge questionnaire	66
Table 4.2	Financial attitude questionnaire	67
Table 4.3	Financial behaviour questionnaire	67
Table 4.4	Training dataset	68

Table 4.5	Comparison results on the RMSE metric	71
Table 4.6	Comparison results on the MAE metric	71
Table 4.7	Aggregate variable importance	72
Figure 4.1	Histogram and summary statistics – predicted financial literacy	74
Table 4.8	Avg. balance and return by financial literacy score	74
Figure 5.1	Monthly customer churn rates	78
Figure 5.2	Super guarantee contribution intervals	79
Figure 5.3	Pseudocode for eCSP implementation	82
Figure 5.4	Account churn by days since last contribution	84
Figure 5.5	Data processing flow	87
Figure 5.6	Prediction framework	87
Table 5.1	Training dataset	88
Table 5.2	Model performance	92
Table 5.3	Sequential pattern results	93
Table 5.4	Variable importance by Gini reduction – top 10	94
Figure 5.7	Attrition score vs. financial literacy score	95

Abstract

Through a combination of product innovation and government reform the retirement savings system in Australia has become increasingly complex. Everyday Australians are now required to possess sophisticated financial decision making skills in order to safely navigate the many risks and opportunity costs associated with their superannuation. The decisions that savers make regarding their investments will have a significant impact on their retirement wellbeing. As a result, it is paramount for any responsible financial institution to actively measure, monitor and elevate the financial literacy of its members. To date, there is no research which proposes a suitable context specific construct for financial literacy in superannuation which is predictive of financial outcomes and utilises passive administrative data to enable ongoing measurement. To address these challenges this research first proposes a measurement construct for financial literacy in superannuation informed by the results of a financial literacy survey enriched with administrative member data. Next, I propose a novel solution for the prediction of superannuation literacy using a vast dataset of demographic and behavioural features. The prediction framework addresses the issue of non-response bias while maximising predictive performance. Finally, the prediction framework is validated against a real world business problem, customer churn. The findings of this research indicate that the measurement construct for superannuation literacy significantly outperforms the conventional measure against a number of financial outcomes. Superannuation literacy outperforms common financial literacy by a multiple of 7.1, 11.2 and 8.9 for account balance, portfolio return and portfolio volatility respectively. The prediction framework for superannuation literacy outperforms a number of state of the art algorithms for prediction. The aggregate measure for superannuation literacy achieves an R square of 84.6% and is highly correlated to positive financial outcomes in super. Validation against customer churn provides an insight into the complex relationship between financial sophistication and decision making. This research provides the framework and tools to monitor and engage superannuation members based on their sophistication and intervene where they are determined to be at-risk, requiring additional support to manage their retirement savings, or to maximise member satisfaction and engagement.

Keywords – Financial literacy, superannuation, passive system, outcomes-based measurement

Table of Contents

1. Introduction 1
   1.1. Motivation 2
   1.2. Stakeholders, Aims, Objectives, and Significance 3
      1.2.1. Stakeholders 3
      1.2.2. Aims 4
      1.2.3. Objectives 5
      1.2.4. Significance 6
   1.3. Research problems 8
   1.4. Research contributions 8
      1.4.1. Financial literacy for superannuation members using survey data 9
      1.4.2. A prediction model for financial literacy in superannuation utilising passive administrative data 10
      1.4.3. Validation of the prediction framework against business outcomes 10
   1.5. Structure of the thesis 11
2. Preliminaries and literature review 12
   2.1. Superannuation in Australia 12
   2.2. Financial literacy measurement and applications 13
      2.2.1. Financial literacy and well-being 13
      2.2.2. Defining financial literacy 14
      2.2.3. Measurement of financial literacy 16
      2.2.4. Financial behaviour 19
      2.2.5. Financial attitude 20
      2.2.6. Explanatory variables for financial literacy 21
   2.3. Machine learning techniques for measuring financial literacy 22
      2.3.1. Classification and regression algorithms 23
      2.3.2. Handling class imbalance 25
      2.3.3. Evaluation metrics 27
   2.4. Summary 29
3. Financial literacy in superannuation: Empirical evidence for positive financial outcomes and decision making 31
   3.1. Introduction 31
   3.2. Related work 33
      3.2.1. Measurement of financial literacy 33

3.2.2. Outcomes of financial literacy 34
3.3. Methodology 36
3.4. Survey results 36
3.4.1. Basic financial knowledge 37
3.4.2. Superannuation knowledge 41
3.4.3. Financial attitude 46
3.4.4. Financial behaviour 48
3.4.5. Self-assessed financial literacy 50
3.5. Association to positive outcomes 52
3.5.1. Risk, return and account balance 52
3.5.2. Saving behaviour and tax effective contributions 55
3.5.3. Number of investment options 57
3.6. Conclusion 58
4. Predicting financial literacy in superannuation using debiased multi-output regression 60
4.1. Introduction 60
4.2. Related work 62
4.2.1. Prediction of financial literacy 62
4.2.2. Multi-output Regression 63
4.3. Prediction framework 64
4.3.1. Problem definition 64
4.3.2. Identifying non-response bias 65
4.3.3. Chained multi-output regression 65
4.4. Experiment 65
4.4.1. Data collection 66
4.4.2. Experimental setup 69
Evaluation Protocols 69
Baselines 69
4.5. Discussions 71
4.6. Conclusion 75
5. Modelling unseen churn using sequential pattern features 76
5.1. Introduction 76
5.2. Related work 80
5.2.1. Customer Churn Prediction 80
5.2.2. Contrast Sequential Pattern Mining 81

5.3. Problem Statement 83
5.4. Prediction Framework 84
5.4.1. Classifier 85
5.4.2. Designing Static and Sequential Features 85
5.5. Results and Discussion 87
5.5.1. Data Collection 87
5.5.2. Evaluation Metrics 90
5.5.3. Prediction Performance 91
5.5.4. Analysis of Feature Importance 93
5.6. Churn and financial sophistication 94
5.7. Conclusion 95
6. Conclusion 97
6.1. Contributions 98
6.2. Applications 99
6.3. Future work 100
Bibliography 101