# http://www2.phys.canterbury.ac.nz/~u-lenser/images/uc_logo_white_transparent.png



|  |
| --- |
|  |
| **Project Summary: Predicting Financial Difficulty and Hardship with Machine Learning at ANZ** |
|  |
| **DATA601 Project 2022**  **Ally Hassell** |
| **University of Canterbury & ANZ Good Customer Outcomes**  **Date Submitted: 10/02/2023** |

# Research goal

ANZ’s Good Customer Outcomes team wanted to explore the use of advanced data analytics improve customer outcomes. **This project used ANZ’s financial data to create a machine learning model to predict customers likely to experience future financial difficulty or hardship.** Customer outcomes could be improved because this information allows ANZ to provide pre-emptive support. This is reflective of the Good Customer Outcomes teams guiding principle 9, which is to provide ANZ customers with support when they need it.

Financial difficulty occurs when there are concerning financial behaviours such as frequent repayments are missed. Hardship means a customer has, or likely to, default.

# Data

Customer habits and behaviours were summarised in a single table, with a row for each individual customer. Each column was measured a feature of the customers financial patterns, such as the percentage of their income spent on necessities. New features were created and measured to improve the model and maximise the benefit to ANZ. This step was completed using SAS code.

The novel features were:

* Spending velocity: what percentage of the time between income payments did it take for a customer to spend 75% of their income.
* Pay later: number and amount spent through pay later schemes (e.g. Afterpay).
* Credit card: proportion of their credit card which was discretionary or necessity spending.
* Expenses: proportion of their income which was discretionary or necessity spending.

Other features from the ANZ Assist program were used:

|  |  |
| --- | --- |
| * Bank categories: credit rating * Credit cards: late fees, number * Financial summaries: FUM | * Overdraft: limit, average balance * Personal loans: amount, amount payed off * Home loans: Number, type, missed payments |

This analysis aimed to capture the change in features over time. Six months of past data was used, going back from either January 2023 or the date a customer entered financial difficulty/hardship.

# Machine learning

Machine learning models work but using algorithms to find the patterns and relationships within data, and then uses this knowledge to make predictions about future events. Two models were created, one for stable/difficulty and stable/hardship. To build a model which was suitable for this project it needed to maximise the number of correct financial difficulty predictions while also reducing the number of false positives. Thirteen models were created and compared in Python.

# Results

Random forests performed the best for financial difficulty and extremely randomized decision trees performed the best for financial hardship. Comparing the performance of financial difficulty and hardship, the significantly lower results for hardship demonstrate that it is more difficult to predict.

The number of correct classifications in relation to the number of false positives and negatives:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  |  | **Actual** | | |  |  | Positive | Negative | | **Predicted** | Positive | True positive | False positive | | Negative | False negative | True negative |   Key | |  |  |  |  | | --- | --- | --- | --- | |  |  | **Actual** | | |  |  | Positive | Negative | | **Predicted** | Positive | 7,153 | 1,049 | | Negative | 5,462 | 1,881,137 |   Financial difficulty (random forest) | |  |  |  |  | | --- | --- | --- | --- | |  |  | **Actual** | | |  |  | Positive | Negative | | **Predicted** | Positive | 47 | 226 | | Negative | 1,452 | 1,894,575 |   Hardship (extremely randomized decision trees) |  |  |

For financial difficulty, the table shows that roughly two third of customers were able to be identified with a relatively low false positive rate. The financial hardship model was only identifying a small proportion of the customers with significantly more relative false negatives.

The features which were important to improving the model’s accuracy were:

|  |  |
| --- | --- |
| Financial difficulty   * Percentage of income on necessities * Median velocity * Number of pay later payments * Daily credit card spending | Hardship   * Percentage of income on necessities * Median velocity * Number of pay later payments * Number of floating home loans |

Original features are denoted with a ( ). Creation of the novel features was able to add to the models performance and enrich its insights. These features could be used as a basis for early warning indicators.

# Next steps

This project has successfully demonstrated a the proof of concept. It requires additional work before it can be implemented, so next steps would be to make the models production ready so that its insights can be used by ANZ’s Good Customer Outcomes team. To be production ready the program requires the use of data science techniques to authenticate and validate its predictions. Hardship models could employ additional data science methods to determine if its performance could be improved to an acceptable level of accuracy.

The models could be improved by incorporating class imbalance techniques such as under or over sampling, which may be particularly useful in improving financial hardship recognition rates. Additionally, the models would benefit if methods such as encoding variables, resampling or model hyperparameter tuning were used. New features could also be added to enrich the dataset and improve the models accuracy. Accommodation is often a customer’s biggest expense, so the proportion of customers income they spend on their home loan or rent could be useful.

Once improved and peer reviewed, the model could be used to make live predictions and provide the Good Customer Outcomes team with leads on customers to contact and support. The financial features important to the model could be used as a tool by the Good Customer Outcomes team to implement evidence-based strategies to prevent these unfavourable financial outcomes. Experiencing financial difficulty and hardship negatively impacts customers and is costly to the bank; these models can provide ANZ with a tool to affect change in this space.