Raising Survey Response Rates by Using ML Method to Predict Gold Providers

Summer Wang
Australian Bureau of Statistics
Setting the Scene

- Resourceful paradata
- Advanced ML techniques
- Limitations of current follow-ups
- Limited resources for follow-up
- Declining response rate

Responsive follow-up framework
A Responsive Follow-up (FU) Framework Using Response Propensity (RP)

First Phase
- Identifying self-responders, AKA., gold provider (GP) units and conduct the GP strategy
  - The non-GPs will be followed up with calls as normal
  - The GP units will not have any FU action being undertaken.
  - The FU resources saved from these GP units will be re-allocated to the non-GPs

Second Phase
- Subgrouping non-respondents and allocating different call-caps

Final Phase
- Identifying potential non-respondents, AKA., black units, with no further FU attempts to be made towards them
Predicting RP

- The influencing factors to predict the RP
  - Survey variables
  - Paradata info

- The ML method to predict the RP
  - The Random Forest of Regression Trees method
    - Parameter tuning, no. of trees, mtry, node size, etc.
    - Data training – the 10-Fold Cross Validation
    - Modelling results – response propensity scores (0-1)
The Rural Environment and Agricultural Commodities Survey (REACS)
- Sample size – around 27k
- Response rate – around 80%

The Live Trial of the GP strategy
- GP status predicted for all continuing sample units
  - Rule based approach
  - The Random Forest of Regression Trees method
- Entire survey sample split into control and treatment groups
  - The control group will receive the standard FU practice
  - The treatment group will be conducted with a GP strategy
Key Findings from the Live Trial

- The GP strategy promoted cost-efficiency for the FU practices.
- The GP strategy posed no additional data quality risk, i.e., estimates and variances.
- The modelling approach outperformed the rule-based approach:
  - More accurate in prediction results.
  - More flexible and adaptive for application.
Current Status

- It has been implemented by the REACS subject matter area (SMA) on an annual basis
- It has also been applied to a few other business surveys with more to be rolled out

Practicality is key

- The compatibility with current infrastructure
- The capability of SMA to self-serve
- The feasibility as a corporate level tool
Partition the sample units by their RP into response homogeneity groups (RHGs)

- For existing sample units, RF method
- For new sample units: kNN imputation method

Apply the Quantitle regression (QR) inverse probability weighting (IPW) estimator using RHGs for estimation

Optimising the allocation of FU resources with the aim to reduce the overall mean squared error of the estimates.

- The key is to find optimal number of follow-up contacts for different RHGs based on their probability of a successful conversion.
Full Details:

1. **The GP Strategy Paper**

2. **The Expanded Resource Allocation Strategy Paper**

Feel free to contact me via summer.wang@abs.gov.au