Predicting the Type of Financial Services Consumer for got moneys?Inc.

by Mariya Petukhova

May 2017

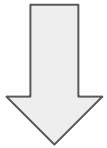
Disclaimer

“At the request of the survivors, the names have been changed. Out of respect for the dead, the rest has been told exactly as it occurred.”

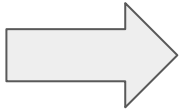
got moneys?Inc. is not real !!!

Client's Challenge

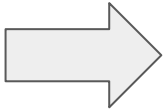
Who are my customers?



New Segments



???



Profit

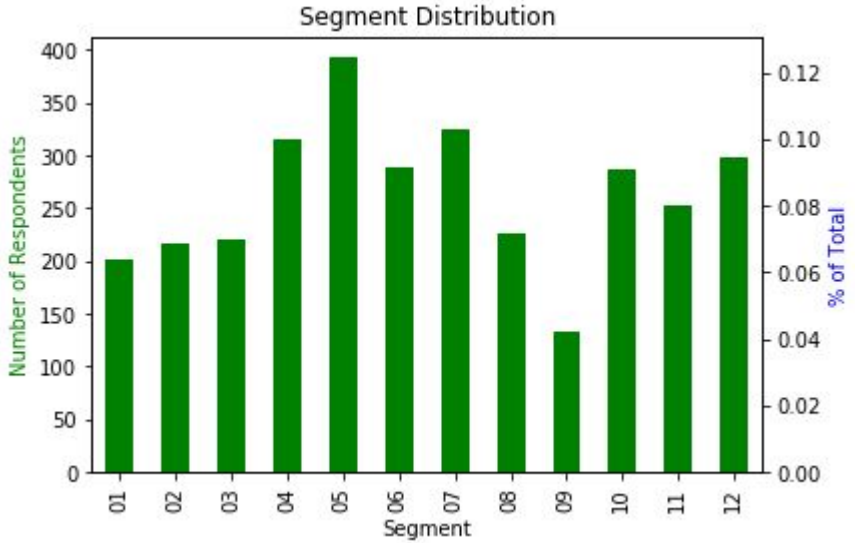
Data Scientist’s Challenge

- Develop an algorithm to classify got moneys?Inc. customers and prospects into newly developed* consumer segments using 3rd party data
  - Segments are limited to survey respondents
    - How to classify new instances
    - Survey sample is small
  - Client’s customer database is incomplete and skewed towards certain lines of business
  - 3rd party data is not great in capturing qualitative measures collected via surveys

* via 2-stage nuanced survey tasked with collecting data on people’s attitudes and enduring values towards personal finances (e.g., risk aversion, optimism, hustle vs flow, financially self-confident)

Data

- 3,154 respondents of Financial Market Structure and Consumer Segmentation Survey
  - Average 263 instances per segment
- 671 features (Acxiom Data Discovery 2 package, recoded*)
- 12 Consumer Segments
  - Thus, average random accuracy is 8.3%



*After further exclusion of scarcely populated features (90% or more missing values) 493 features left for modeling

Top 10 Most Important Features

- Age
- Home Market Value - Estimated
- Home Length of Residence
- Economic Stability Indicator
- UnderBanked Indicator
- Heavy Transactors
- Media Channel Usage - Cell Phone
- Media Channel Usage - Primetime TV
- Brand Name Medicine Propensity Score
- Tele Trends - International Long Distance User

Mostly Socio-economics factors

Solutions

- Random Forest
  - Good for wide data
  - Feature Importance
- Support Vector Machine (with 10-component PCA)
  - Good counterpart to Random Forest
- Multi-layer Perception Classifier (with 10-component PCA)
  - Stochastic gradient-based optimiser

Results

- Best Accuracy is achieved via Random Forest Classifier with max_depth=2 and n_estimation=900
  - 19% accuracy over 8.33% => 228 lift
- Not all segments are predicted equally: the model does a great job at predicting segment 5 (81% accuracy) and can also be used for segments 4 and 12; the rest of segments are not predictable at this point

Next Steps

- Create segmentation code for implementation
- Get paid
- Get segments descriptions and fine-tune feature design accordingly
- Collect more data
    - Consider adding 2 or 3 segment-defying survey questions on applications
    - Test segment-specific marketing campaigns and products to let consumers self-identify

got moneys?

Thank you.