Sprocket Central Pty Ltd

Predicting High Value Customers:

Proposed Analytics Approach

Overview

1

Data Exploration

- Define **metric** measuring **individual customer value**.
- Explore **relationship** of possible **predictor variables** with this metric.

2

Model Development

- Use results of data exploration to **determine which model** is appropriate.
- Shape **customer demographic** data into **training set**.
- **Fit model** and measure in and out of sample **accuracy**.

3

Prediction and Interpretation

- Form **new customer data** into predictor set.
- **Predict customer value** metric using new customer data.
- **Rank customers** based on customer value metric to determine **high value customers**.

Data Exploration

Customer Value Metric: In order to **determine high value customers**, we would like to be able to **predict a metric** that measures any particular customer's **value to Sprocket** over a given time period - for example **profit generated in a year**.

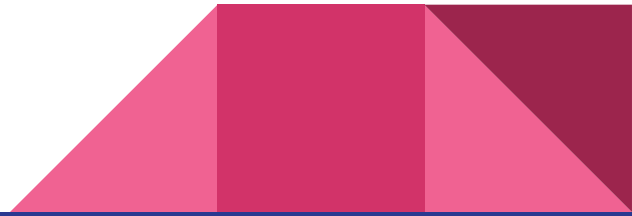
Calculating the Metric: We can calculate this for each customer by grouping their transactions by their customer ID in the transactions data set, then subtracting the **summed standard costs** from the **summed list prices** across all their purchases to find the **profit** they have generated. Then we can look at the time period over which these transactions were made and **normalise** to some **standard period of time**, say a day or a year.

Data Exploration Continued

Explore Relationship of Metric with Predictor Variables

Determine which attributes may be good **predictors** of the profitability metric and examine the **pairwise relationships** between each predictor variable and the dependent variable, profit per year, looking for linearity, correlation or any other kind of relationship.

This can be done **visually** using plots, or mathematically using **hypothesis tests** and **correlation coefficients** if necessary.



Model Development

Choose Model: Use the information gained about the **relationship** of possible predictor variables with the customer profitability metric to **determine what sort of model** to use. For example, if the relationships are largely linear, use multiple regression. If not, use tree based regression methods.

Fit Model: Once the model is chosen, use the **customer demographic** test set to **fit the model** to predict customer profit per year.

Measure Accuracy: Use a **train-test split** or **cross-validation** to measure **in and out of sample accuracy** using metrics such as **Root Mean Squared Error** (RMSE) and **Mean Absolute Error** (MAE).

Prediction and Interpretation

Predict Profitability of Each Customer: Adapt the new customer list data into a form identical to that of the training set - that is a data frame with the same predictor variables as columns. Then run the **predict** function.

Determine High Value Customers: Once the vector of predictions has been obtained, attach these figures to the appropriate rows of the new customer list data set. Then **sort the customers** by their predicted profit - highest to lowest.

Choose some **threshold of profitability** and determine which customers exceed that threshold.These are the customers that should be **targeted** as being **high value**.

Thanks!

Contact: **Ben Crabtree**

- bcra4120@uni.sydney.edu.au
- https://www.linkedin.com/in/benjamin-crabtree/