Acquisition Modeling Analysis

Michael Steck

Davenport University

Acquisition Modeling Analysis

The acquisition modeling technique chosen for this analysis was the cross-selling model. This model seeks to sell additional products to existing customers by identifying the data patterns that estimate the targets propensities for purchasing the product. This is done by using historic customer data, where the customers have not previously purchased the product, to train the model. Customers with this same profile can then be appropriately targeted in the campaign for the cross-sell opportunities. (Chorianopoulos, 2016, pp. 97 - 98)

For business understanding, this analysis was based off the goal to increase sales of products in the movie product category by cross-selling. The transaction dataset is assumed to be a subset of transactions where the customers had not previously purchased movie products. The resulting model will help the business understand the data factors and patterns that indicate a customer’s propensity to buy movies.

The first step in this process is to consider and explore the three datasets that are available for modeling. These datasets include a transaction, a promotion, and a product category file. The three files were individually inspected for data understanding and data quality. Based on this analysis, it was determined that the three files could be cleanly merged into one combined dataset. This included the creation of two new data fields that normalize the field values of unit price and sales amount.

The next step in the data preparation was to consider the product categories field. This field indicated the product category for each transactional sale. However, this information was stored in one field making it difficult to use in the upcoming modeling activities. This field needed to be restructured to where every product category was its own field. A simple flag indicator, in the newly created fields, could now show the category of the sold product. While these new flag fields could work well in the modeling, they could be strengthened by identifying the quantity of items purchased. This resulted in fields that showed the quantities of customer purchased items.

This now presented another challenge. The quantity of products purchased in the smallest set ranged from 0 to 3 and the largest set from 0 to 75. A range varying this much could lower their predictive capabilities. This was resolved by deriving a new normalized set of fields that represented a ratio, or proportion, of the product sale. As a result, the values now ranged from 0 to 1 for all product categories.

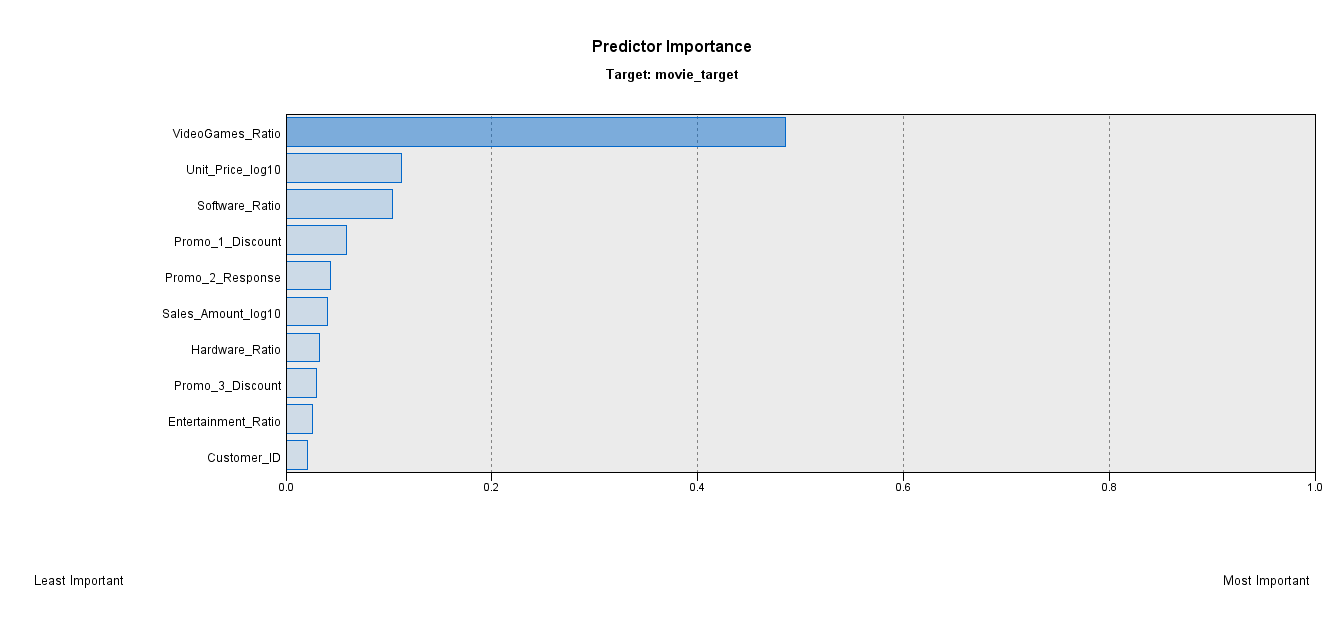
Not to be forgotten, a movie target field was added prior to the product category normalization. The movie product category field did not need to be normalized, but converted to a flag. If the sum of the customer purchases in the movie category was greater than 0 then a 1 was placed in the movie target field. Otherwise, a 0 was placed in the movie target field indicating that there were no sales in this category.

Final data preparation included a check for perfect predictors and adding a data partition to the dataset for training and testing. It should also be noted that temporary working, correlated, and unwanted fields were removed prior to modeling.

Modeling began with the Auto Classifier set to return six models. Of these six models, the C5 stood out with an initial accuracy score over 92%. This was over 20% more accurate than the next top performer XGBoost Tree at roughly 71%. As a result, modeling continued with a focus on tuning the C5 model to optimize performance. The resulting boosted C5 model performed extremely well as shown in the following table.

|  |  |
| --- | --- |
| Gain | Lift |
| ROC | Confusion Matrix |

The boosted model did perform slightly better against the testing dataset when compared to the non-boosted, by roughly 1%. While the boosted model did perform better, the non-boosted model more clearly showed the importance of certain factors. In this case, purchasing a product in the Video Game category made the customer much more likely to buy a movie. The next two importance factors included Unit Price and the Software category purchases.



In conclusion, a strong classification model was created to aid in cross-selling campaigns. Key findings included the discovery of a link between Movie sales and Video Game sales. Also, the importance of unit price should be noted. This factor could point to some customer price sensitives, but exploring this was beyond the scope of this analysis.

# References

Chorianopoulos, A. (2016). *Effective CRM Using Predictive Analytics.* West Sussex, United Kingdom: John Wiley & Sons, Ltd.