Insurance Cross-Sell Prediction

Ali Malenchik

DSC 680 T301-2221 Fall 2021

https://github.com/alimalenchik/Portfolio

Which Domain?

This data comes from the insurance sales domain.

References:

- https://www.investopedia.com/terms/c/cross-sell.asp
    - A logical definition of cross-selling, how it works in different industries, its advantages & disadvantages, and real-world examples.
- https://www.analyticsvidhya.com/blog/2021/06/cross-sell-prediction-solution-to-analytics-vidya/
    - A sample machine learning project which explores one approach to the cross sell prediction problem.
- https://medium.com/@opialabs/cross-sell-prediction-part-1-of-3-a9a6bcc4e5a8
    - Another sample machine learning project which explores one approach to the cross sell prediction problem.
- https://www.kaggle.com/priyankaparlikar/target-customer-for-cross-selling-of-bank-product
    - Another sample machine learning project which explores one approach to the cross sell prediction problem.
- https://medium.com/analytics-vidhya/finding-cross-sell-opportunities-with-machine-learning-d47669b36f68
    - An introduction to how machine learning is used in cross-selling.

- https://www.ibm.com/blogs/nordic-msp/six-ways-use-data-science-drive-cross-sell-upsell-activity/
    - An overview of cross-selling tasks in machine learning, including customer segmentation, uplift modeling, product recommendation, product association analysis, clustering, and neural networks.
- https://www.netguru.com/blog/machine-learning-sales-cross-sell-upsell-campaigns
    - A guide to real-world cross-selling and upselling strategies along with their purpose.
- https://addepto.com/up-selling-cross-selling-5-reasons-use-machine-learning/
    - Details the benefits of using machine learning for cross-selling and sales prediction.
- https://algonomy.com/blogs/personalization/solving-the-limitations-of-traditional-cross-sell-models-with-deep-learning-and-nlp/
    - A deep dive into how neural networks create cross-sell recommendations.
- https://www.salesforce.com/eu/learning-centre/sales/cross-selling/
    - A step by step guide for cross-selling to customers, including examples, benefits & drawbacks, and techniques.

Which Data?

The dataset I will be examining is a collection of information on cross-sale outcomes for 381,109 health insurance policyholders found on Kaggle. The dataset is provided in CSV format and contains twelve columns:

- ID
- Gender
- Age
- Driving_License

- Region_Code
- Previously_Insured
    - 1: Customer already has vehicle insurance
    - 0: Customer does not have vehicle insurance
- Vehicle_Age
- Vehicle_Damage
    - 1: Customer’s vehicle has been damaged in the past
    - 0: Customer’s vehicle has not been damaged in the past
- Annual_Premium
- Policy_Sales_Channel
    - Anonymized code for the method of outreach (i.e. different agents, mail, phone, in person, etc.)
- Vintage
    - Number of days the customer has been associated with the company
- Response
    - 1: Customer is interested in vehicle insurance
    - 0: Customer is not interested in vehicle insurance

The target variable for prediction is “Response”, which indicates whether the health insurance policyholder responded as interested in vehicle insurance from the same company.

Research Questions? Benefits? Why analyze these data?

Cross-selling is the concept of marketing additional products to existing customers in order to earn additional revenue. Cross-selling is a common practice in financial industries, insurance industries, and more. Using machine learning to predict which customers are responsive to cross-selling allows businesses to effectively target communications and optimize marketing strategies.

My research questions are as follows:

- What is the distribution of customer response in the dataset?
- What trends can be observed for each independent variable and its relationship with customer response?
- Is there any correlation between independent variables?
- Which features are important for prediction?
- Which model is best performing? How can the model be optimized and to what extent?

What Method?

To begin, the data will be split into a test and train dataset in order to reserve unseen data for model testing and evaluation. Exploratory data analysis will be used to visualize trends and relationships between the variables. This will include multiple bar and scatter plots, histograms, and correlation analysis.

Once exploratory data analysis is complete, categorical features will be encoded to prepare for modeling. Feature selection will be performed using sklearn's mutual_info_classif() function. The most important features will be used to create multiple classification models, including Random Forest, Logistic Regression, and Naïve Bayes. Models will be evaluated based on multiple hyperparameter values to determine the best performing model.

Potential Issues?

I expect the hyperparameter tuning to be the most time consuming and challenging part of this project. I also foresee there may be potential issues when encoding the data; this may result in a large number of features which could impact model performance. Additionally, because the dataset is so large the model training and evaluation may be time consuming.

Concluding Remarks

Cross-selling is a valuable tool that allows businesses to market products to an already existing client base. Using machine learning to predict which customers are responsive to cross-selling is more efficient and cost-effective, reducing marketing waste and increasing revenue. Because historical information can be gathered on the responsiveness of customers along with other customer characteristics, cross-selling is a prime candidate for training a classification model to predict success of a cross-selling attempt.