CNA

Predictive Modeling Best Practices

Chuck Boucek
VP and Actuary
Predictive Modeling
CNA Insurance

CNA

Project Life Cycle

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Model evaluation
- Deployment
- Monitoring
- Post Deployment Support

CNA

Required Project Roles

- Data Analyst
- Modeler
- Pricing Actuary
- Underwriting Champion
- Senior Staff Champion
- Claims professional

CNA

Predictive Modeling Best Practices

Business Understanding

- Developing a detailed knowledge of the current process
- SWOT analysis
- Gather initial hypotheses regarding expected patterns in the data

The business understanding phase should end with a clearly stated goal for the final deployed model

3

CNA

Predictive Modeling Best Practices

Data Understanding

- Collect the needed data in its raw format
- Explore data
  - Conduct initial high level univariate analyses in order to determine how the data can best be employed
- Examine data quality

The data understanding phase concludes with the construction of documentation that describes

1) The data collected

2) Conclusions regarding how the data is to be employed

3) How any data quality concerns will be addressed

4

CNA

Predictive Modeling Best Practices

Data Preparation

- Never underestimate the amount of time required for data preparation
- The data preparation process should serve five main functions. These different data functions may be served by a single data source.
  - Extract and prepare the data for model construction
  - Extract and prepare the data for model testing
  - Extract and prepare the data for performing impact analysis
  - Extract and prepare the data for testing the implementation of the model in internal systems.
  - The data needs for monitoring the results following deployment of the model must be considered.

5

CNA

Predictive Modeling Best Practices

Data Preparation (continued)

- Construction of appropriate check values
- Missing values – proper treatment of missing values of must be considered.
  - Common Options for treating missing data
    - Deletion of records with missing data
    - Mean imputation
    - Mean imputation with missing value indicator
    - Imputation algorithm
  - Other pertinent questions
    - How will missing data be treated at time of implementation?
    - How to treat "sort of" missing data?

6

CNA

Predictive Modeling Best Practices

Data Preparation (continued)

- Loss development at the individual claim level
  - State all claims at a common development period
  - Develop only open claims
  - Separate development by type of claim
  - Separate development by class of business
  - Model of ultimate claim values
- Loss Trend
- Loss Capping
- Treatment of different deductibles
- Catastrophe losses – generally an underwriting model will exclude cats
- Benefit and Coverage level changes

7

CNA

Predictive Modeling Best Practices

Data Preparation (continued)

- Premium at present rates
  - Adjusting premium to present rates may be required.
  - It becomes a less critical factor if rate action is mostly to maintain rate adequacy against loss trend.
  - It is a more critical factor if significant rate action has been taken above or below historical trend.
  - Since the data in the model is at the individual policyholder level or below, the need and ability to determine on level premium at that level must addressed.

8

CNA

Predictive Modeling
Best Practices

Data Preparation (continued)

- The predictor variables to be employed should be thoughtfully constructed
- Final step is to test database against check values

The data preparation phase concludes with the construction of documentation that contains

1) The location of the input files used in modeling database construction
2) The code used to cleanse the data and create the modeling database
3) The location of the final modeling database
4) A conceptual description of how the technical aspects were addressed

9

CNA

Predictive Modeling
Best Practices

Model Construction

- Model should be constructed at the lowest level of detail that is feasible
- Regular meetings with pricing actuaries, underwriters, IT and senior management throughout the model development process
- Throughout the model development process, reasonability tests should be applied to predictor variables
  - Time consistency
  - cross validation tests

10

CNA

Predictive Modeling
Best Practices

Time Consistency Plot

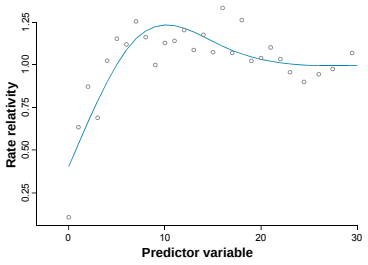
- Partial plots are a key tool to visualize predictor variables throughout the model building process
- What is a partial plot?
  - Linear predictor = $k + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
  - Predicted value = $(e^k) \times (e^{\beta_1 X_1}) \times (e^{\beta_2 X_2}) \times (e^{\beta_3 X_3}) \times (e^{\beta_4 X_4})$
- Partial plot demonstrates an individual predictor variable's contribution to final prediction

11

CNA

Predictive Modeling
Best Practices

Sample Partial Plot

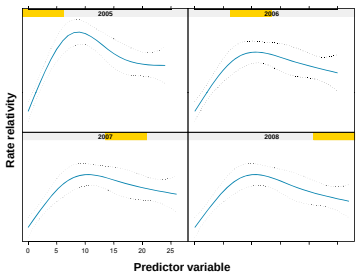


12

CNA

Predictive Modeling
Best Practices

Time Consistency Plot

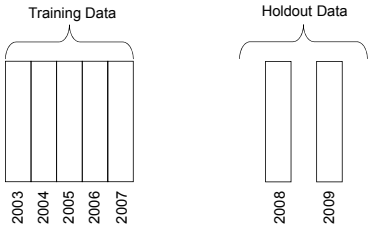


13

CNA

Predictive Modeling
Best Practices

Cross Validation Process



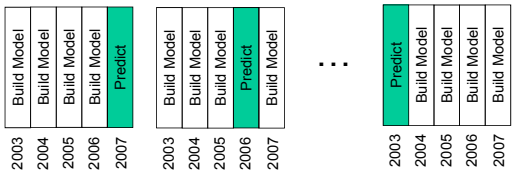
- Predictive modeling process employs training and test data
- Testing process mimics actual model deployment

14

CNA

Predictive Modeling Best Practices

Cross Validation Process



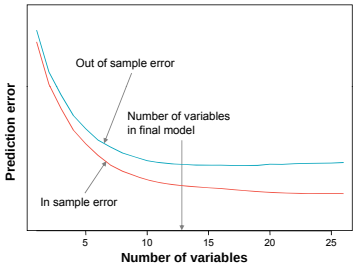
- Cross validation process mimics testing process within the training data

15

CNA

Predictive Modeling Best Practices

Classic Cross Validation Exhibit

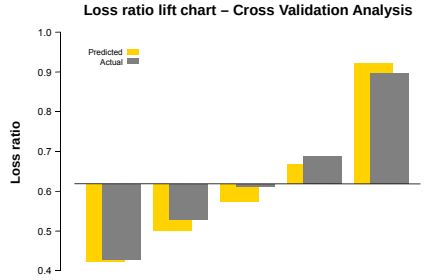


16

CNA

Predictive Modeling Best Practices

Cross Validated Lift Chart



17

CNA

Predictive Modeling
Best Practices

Model Construction

- Tradeoff between variables that are intuitive to underwriters and variables that have an unexpected relationship to loss
  - High touch vs. high tech implementation
- Occam's Razor – "If in doubt, leave it out"
- Credibility in multi-variate modeling is a burgeoning area
- Holdout Sample testing

The Model Construction phase concludes with a report documenting
1) The predictor variables and coefficients in the final model
2) The key tests of statistical significance and visualizations employed for individual predictor variables
3) The key tests of overall model performance

18

CNA

Predictive Modeling
Best Practices

Model Evaluation

- Impact analysis
  - The proposed deployment vehicle should be prototyped in great detail with input from pricing, underwriting and IT
  - All aspects of the new pricing and underwriting process should be documented and then analysis should be constructed to understand how the revised process will impact policyholders.

19

CNA

Predictive Modeling
Best Practices

Model Evaluation

- A final presentation of the model should be made to key constituents for final sign off.
  - One of the key purposes of this presentation is to assess if the initial modeling goals have been met
- In addition to sign off regarding the business use of the model, sign off from the Law Department must also be obtained.

At the conclusion of the model evaluation phase, a report is constructed that
1) Demonstrates that the modeling goals have been met
2) Documents the signoff of the key stakeholders
3) Summarizes the impact analysis and deployment vehicle prototype

20

CNA

Predictive Modeling Best Practices

Deployment

- Data construction must support model deployment
- Clearly defined and rigorously enforced approval process
- Data to support deployment testing
  - Modeling data should mirror the deployment process
  - Data Model

Source → Stage → Input → Reference → Score

At the conclusion of the deployment phase the rubber has met the road

21

CNA

Predictive Modeling Best Practices

Model Monitoring

- Data construction must support ongoing model monitoring
- Is the model having its intended impact in the marketplace?
  - New Business writing
  - Retention
  - Expense Ratio
  - Book of business quality
  - Loss Ratio by model decile
  - Overall loss ratio of book of business

22

CNA

Predictive Modeling Best Practices

Post Deployment Support

- Data construction must support the questions that arise after deployment
- Asses potential IT errors
- Respond to questions from underwriters regarding specific policy scores
  - These questions give a window into model elements causing underwriter angst
    - Is more training needed?
    - Do models need to be revised?

23