

11/28/23

NAIC Rate Model Review Team’s GLM Checklist

Regulators frequently using the NAIC rate model review service asked the NAIC rate model review team to create a list of rate filing documentation needed for the NAIC to complete a full-scope rate model review. The goals of such a list are to make the NAIC review process more efficient and expeditious. Regulators may evaluate this list and determine the state’s needs. Regulators can share this list with insurers, can revise the state’s rate filing checklist, or can communicate with insurers through rate filing objections, when needed.

The list is divided by “Essential Information” and “Sometimes Needed Information.” These terms are defined in this table:

Category	Description
Essential Information	Information that the NAIC rate model review team requests before writing a full-scope initial assessment of a model.
Sometimes Needed Information	Information that the NAIC model review team finds useful for model reviews but may only be needed if something appears non-standard about the modeling approach. Regulators may want to wait to request such information from insurers only when requested in the initial NAIC report.

Model Introduction – Essential Information

- A narrative discussing what the company is trying to accomplish with the model, including the following details:
 - Is this a new model or refresh? What is the prior model’s SERFF number (if applicable)?
 - Does the filing impact existing renewals?
 - Who is the target consumer?
 - What is the GLM intended to model? (Frequency, Severity, Loss Ratios, Pure Premium, etc.)
 - What is being optimized? Does the model consider anything other than differences in loss cost?
- A narrative discussing the specifications and high-level assumptions of the model, including the following details:
 - Number & Type of models (GLM, GBM, etc.)
 - Split of the data into models (by coverage, by peril, etc.)
 - Split of the data into datasets (training, test, holdout)
 - How models were combined to derive the final rating algorithm

Model Introduction – Sometimes Needed Information

- A narrative discussing the credentials of the modeling team, including the following details:
 - Name of each individual
 - Relevant educational experience
 - Relevant credentials and designations
 - Years of experience building predictive models
 - Years of experience in the insurance industry

- Discuss how Actuarial Standards of Practice (ASOPs) 12, 23, 41, and 56 were considered in building the models.
- Describe the software used to build the models.
- Provide copies of or links to academic references for their modeling techniques.
- A table listing the states where the model has been filed for review, the SERFF tracking number, and an indicator showing whether the filing has been approved.

Data – Essential Information

- A narrative providing the description of each data source including the following:
 - Informational materials or website links for each 3rd party
 - Commentary on how the company reviewed the veracity of the data source
 - Why the company believes the data source is useful for the model’s intended purpose
 - Disclosure of known data errors
 - SERFF filing numbers where the use of the data was previously approved (if applicable)
- A description of the relevance of the data
 - The lines of business and companies included should be identified
 - Description of any considerations or adjustments made to make the data more applicable for its intended use
- A data dictionary provided as a table with the following columns:
 - Data Source (Vendor name or “Internal”)
 - Variable name
 - Alternate names appearing in other filing documents
 - Data types (discrete, continuous, logical, categorical)
 - Treatment Type (Model, Control, Offset, Target)
 - Possible values (Empirical min and max for numerical variables, all categories for categorical variables)
- Tables showing summary metrics for each dataset by year (training, testing, holdout)
 - Year
 - Losses
 - Exposures (or Policy Count)
 - Claim Count (if applicable)
- A narrative on how the company determined the final variables to include in the final model
- A narrative on the data accuracy and data reconciliation process
 - Description of the methods used to compile, filter, and/or merge data from different sources
 - How the data was reconciled to other sources
- A listing of the rational explanation for each modeled variable that discusses why it would plausibly impact insurance risk as discussed in the CASTF white paper¹.
- A guarantee that the modeling dataset will be retained for at least 7 years
- A description of any dimensionality reduction techniques (PCA, clustering, etc.) that were applied to the data.
- An Excel file with 100 anonymized sample modeling records including all predictor variables and target variables.

Data – Sometimes Needed Information

- A description of steps taken to meet state requirements regarding unfair discrimination (if applicable).

¹ https://content.naic.org/sites/default/files/inline-files/9-15%20CASTF%20-%20Predictive%20Model%20White%20Paper%209-09-2020_1.pdf

- A listing of variables which are subject to the fair credit reporting act (if applicable).
- A table showing the data volume distribution by state for each dataset (training, testing, holdout)

Modeling – Essential Information

- A narrative discussing the specifications and assumptions of the model, including the following details:
 - Form of the regression equation
 - Distribution assumed for the error term
 - The link function (if applicable)
 - Weights used in regression (if applicable)
 - Hyperparameter values and tuning procedure (if applicable)
- A description of how the model differs from prior versions of the model (if applicable).
- A description for each control or offset variable of why it was necessary to treat them as control/offset variables.
- A description of how the variables with null or missing values will be treated, including the following:
 - A table showing the rate of null or missing values for each variable
 - A description of the scenarios which generated null or missing values
 - A description of how each null or missing value is treated (might include imputation method or simply left in as a control)
 - A description of what happens to null and/or missing values when generated in production. (Is there a rating factor applied for null/missing or is the data populated before policy issuance?)
- A description of any large loss capping applicable to the dataset
 - Identify the size of the large loss cap
 - Identify the percentile of claim severity represented by large loss cap
- A description of adjustments and modifications to the data including trending, loss development, capping at minimums or maximums, and removal of outliers.
- A description of variable transformations applied to the data. The description should include the name of each transformation technique used and an example transformation complete with a sample unadjusted value and a final transformed value.
- A description of each feature engineered variables. The description should include the rationale behind the feature engineered variable and a sample calculation including unadjusted original variable values and the final feature engineered variable value.
- A description of how binning was applied to numeric variables and how categorical variable values were grouped together.

Modeling – Sometimes Needed Information

- Deviance residual plots for each model demonstrating the appropriateness of the model assumptions.

Validation – Essential Information

- A narrative on how the model was validated and assessed for model stability
- A narrative on how the model was assessed for improvement over the prior version of the model (if applicable)
- An Excel file containing model output in this format:
 - Each model is a separate worksheet
 - Column A is Variable Name
 - Column B is Variable Level Name
 - Column C is the coefficient
 - Column D is the p-value (if applicable)
 - Column E is the 95th confidence interval lower bound (if applicable)
 - Column F is the 95th confidence interval upper bound (if applicable)
- Ventile plots (quantile plots with at least 20 buckets) for both state specific data and countrywide data, built on data not used for model training. Each plot should include lines for both predicted averages and actual average.
- Lorenz curve for each model built on countrywide data. The plot should include the Lorenz curve and the equality reference line. The plot should also include the Gini value for the model.
- An Excel file containing correlation matrices in this format:
 - Each model's correlation matrix is a separate worksheet
 - Row 1 and Column 1 include variable names
 - The rest of the table displays the correlation metrics
- Commentary on which correlation metric (Pearson's, Cramer's V, etc.) was provided in the correlation matrix Excel file

Validation – Sometimes Needed Information

- A description of how often the model will be validated against new data in the future
- A double lift chart comparing the newly proposed model and the current model (if applicable)
- Actual vs. Expected plots by model and variable (aka "Univariate Plots") which show the closeness between actual averages and predicted averages.
- AIC tests showing the new AIC after the introduction of each additional predictor variable
- F-nested model tests comparing the full model to subset models excluding one variable at a time to demonstrate the significance of each term. Each test should include the following:
 - F-statistic
 - F-test critical values
 - Numerator degrees of freedom
 - Denominator degrees of freedom
- Variance Inflation Factors (VIFs) for each variable

Implementation – Essential Information

- A description of how the models being filed are ultimately integrated into the company's final rating algorithm
- A narrative about all post modeling adjustments, such as smoothing, mapping to scores, and tempering of factors
- A narrative identifying the variables where deviations from indicated were made and commentary on the reason for the deviations
- A dislocation analysis, including the following:
 - Histograms showing percentage premium change on uncapped and capped basis (if applicable), using buckets of 5%
 - Descriptions of the scenarios with the highest increases
 - Descriptions of the scenarios with the biggest decreases
- Commentary on the differences between rating new and existing policyholders
- An Excel file which documents deviations between indicated and selected in this format:
 - Each model is a separate worksheet
 - Column A is Variable Name
 - Column B is Variable Level Name
 - Column C is the Current Factor (if applicable)
 - Column D is the Indicated Factor
 - Column E is the Proposed Factor
 - Column F is the percentage difference between indicated and proposed. If the absolute value of the percentage difference is > 10%, the cell should be highlighted.
- Sample rating/scoring exhibits for 10 risks in Excel, which show risk characteristics, all intermediate adjustments, and the final algorithm output considering the company's final selections.

Implementation – Sometimes Needed Information

None are listed at this time.