APPENDIX B-TREES – INFORMATION ELEMENTS AND GUIDANCE FOR A REGULATOR TO MEET BEST PRACTICES' OBJECTIVES (WHEN REVIEWING TREE-BASED MODELS)

This appendix identifies the information a state insurance regulator may need to review a Tree-based predictive model used by an insurer to support a personal automobile or home insurance rating plan. Tree-based predictive models include Random Forest (RF) and Gradient Boosting Machines (GBM). The list of information elements below is lengthy but not exhaustive. It is not intended to limit the authority of a regulator to request additional information in support of the model or filed rating plan. Nor is every item on the list intended to be a requirement for every filing. However, the items listed should help guide a regulator to sufficient information that helps determine if the rating plan meets state-specific filing and legal requirements. Documentation of the design and operational details of the model will help ensure the business continuity and transparency of the models used. Documentation should be sufficiently detailed and complete to enable a qualified third party to form a sound judgment on the suitability of the model for the intended purpose. The theory, assumptions, methodologies, software, and empirical bases should be explained, as well as the data used in developing and implementing the model. Relevant testing and ongoing performance testing need to be documented. Key model limitations and overrides need to be pointed out so that stakeholders understand the circumstances under which the model does not work effectively. End-user documentation should be provided and key reports using the model results described. Major changes to the model need to be documented and shared with regulators in a timely and appropriate manner. Information technology (IT) controls should be in place, such as a record of versions, change control, and access to the model.[1]

Many information elements listed below are probably confidential, proprietary, or trade secret and should be treated as such, in accordance with state laws and/or regulations. Regulators should be aware of their state laws and/or regulations on confidentiality when requesting data from insurers that may be proprietary or trade secret. For example, some proprietary models may have contractual terms (with the insurer) that prevent disclosure to the public. Without clear necessity, exposing this data to additional dissemination may compromise the model's protection.[2] Although the list of information is long, the insurer should already have internal documentation on the model for more than half of the information listed. The remaining items on the list require either minimal analysis (approximately 25%) or deeper analysis to generate for a regulator (approximately 25%).

The "Level of Importance to the Regulator's Review" is a ranking of information a regulator may need to review, which is based on the following level criteria:

Level 1 – This information is necessary to begin the review of a predictive model. These data elements pertain to basic information about the type and structure of the model, the data and variables used, the assumptions made, and the goodness of fit. Ideally, this information would be included in the filing documentation with the initial submission of a filing made based on a predictive model.

Level 2 – This information is necessary to continue the review of all but the most basic models, such as those based only on the filer`s internal data and only including variables that are in the filed rating plan. These data elements provide more detailed information about the model and address questions arising from review of the information in Level 1. Insurers concerned with speed to market may also want to include this information in the filing documentation.

Level 3 – This information is necessary to continue the review of a model where concerns have been raised and not resolved based on review of the information in Level 1 and Level 2. These data elements address even more detailed aspects of the model. This information does not necessarily need to be included with the initial submission, unless specifically requested by a particular state, as it is typically requested only if the reviewer has concerns that the model may not comply with state laws and/or regulations.

Level 4 – This information is necessary to continue the review of a model where concerns have been raised and not resolved based on the information in Level 1, Level 2, and Level 3. This most granular level of detail is addressing the

[1] Bourdeau, M., 2016. "Model Risk Management: An Overview," The Modeling Platform, Issue 4, December. Accessed online at https://www.soa.org/globalassets/assets/library/newsletters/the-modeling-platform/2016/december/mp-2016-iss4.pdf
[2] There are some models that are made public by the vendor and would not result in a hindrance of the model's protection.

basic building blocks of the model and does not necessarily need to be included by the filer with the initial submission, unless specifically requested by a particular state. It is typically requested only if the reviewer has serious concerns that the model may produce rates or rating factors that are excessive, inadequate, and/or unfairly discriminatory.

Appendix B-TREES is focused on Tree-based models including RFs and GBMs. This appendix should not be referenced in the review of other model types. Tree-based approaches have many significant differences from GLMs. This Appendix B-TREES is intended to provide state guidance for the review of rate filings based on Tree-based models.

A. SELECTING MODEL INPUT

Section	Information Element	Level of Importance to the Regulator's Review	Comments
1. Available Data Sources			
A.1.a	Review the details of sources for both insurance and non-insurance data used as input to the model (only need sources for filed input characteristics included in the filed model).	1	Request details of data sources, whether internal to the company or from external sources. For insurance experience (policy or claim), determine whether data is aggregated by calendar, accident, fiscal, or policy year and when it was last evaluated. For each data source, get a list of all data elements used as input to the model that came from that source. For insurance data, get a list all companies whose data is included inthe datasets. Request details of any non-insurance data used (customer-provided or other), whether the data was collected by use of a questionnaire/checklist, whether data was voluntarily reported by the applicant, and whether any of the data is subject to the federal Fair Credit Reporting Act (FCRA). If the data is from an outside source, find out what steps were taken to verify the data was accurate, complete, and unbiased in terms of a relevant and representative time frame, representative of potential exposures, and lacking in obvious correlation to protected classes. Note: Reviewing source details should not make a difference when the model is new or refreshed; refreshed models would report the prior version list with the incremental changes due to the refresh.
A.1.b	Reconcile aggregated insurance data underlying the model with available external insurance reports.	4	Accuracy of insurance data should be reviewed. It is assumed that the data in the insurer's data banks is subject to routine internal company audits and reconciliation. "Aggregated data" is straight from the insurer's data banks without further modification (i.e., not scrubbed or transformed for the purposes of modeling). In other words, the data would not have been specifically modified for the purpose of model building. The company should provide some form of reasonability check that the data makes sense when checked against other audited sources.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.1.c	Review the geographic scope and geographic exposure distribution of the raw data for relevance to the state where the model is filed.	2	Many models are developed using a countrywide or a regional dataset. The company should explain how the data used to build the model makes sense for a specific state. The regulator should inquire which states were included in the data underlying the model build, testing, and validation. The company should explain why any states were excluded from the countrywide data. The company should provide an explanation where the data came from geographically and that it is a good representation for a state; i.e., the distribution by state should not introduce a geographic bias. However, there could be a bias by peril or wind-resistant building codes. Evaluate whether the data is relevant to the loss potential for which it is being used. For example, verify that hurricane data is only used where hurricanes can occur. The company should provide a demonstration that the model fits well on the specific state or surrounding region.
2. Sub-Models			
A.2.a	Consider the relevance of (i.e., whether there is bias) of overlapping data or variables used in the model and sub-models.	3	Check if the same variables/datasets were used in the model, a sub-model, or as stand-alone rating characteristics. Tree-based models handle redundant variables by splitting on only one of the variables within each component tree. By contrast, generalized linear models (GLMs) struggle with redundant variables as they try to include redundant variables simultaneously. However, best actuarial practice is to keep models as parsimonious as possible and only include additional variables that contribute significant additional predictive power.
A.2.b	Determine if the sub-model was previously approved (or accepted) by the regulatory agency.	1	If the sub-model was previously approved/accepted, that may reduce the extent of the sub-model's review. If approved, obtain the tracking number(s) (e.g., state, System for Electronic Rates & Forms Filing [SERFF]) and verify when and if it was the same model currently under review. Note: A previous approval does not necessarily confer a guarantee of ongoing approval; e.g., when statutes and/or regulations have changed or if a model's indications have been undermined by subsequent empirical experience. However, knowing whether a model has been previously approved can help focus the regulator's efforts and determine whether the prior decision needs to be revisited. In some circumstances, direct dialogue with the vendor could be quicker and more useful.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.2.c	Determine if the sub-model output was used as input to the Tree-based Model; obtain the vendor name, as well as the name and version of the sub-model.	1	To accelerate the review of the filing, it may be desirable to request (from the company) the name and contact information for a vendor representative. The company should provide the name of the third-party vendor and a contact in the event the regulator has questions. The "contact" can be an intermediary at the insurer (e.g., a filing specialist), who can place the regulator in direct contact with a subject matter expert (SME) at the vendor. Examples of such sub-models include credit/financial scoring algorithms and household composite score models. Sub-models can be evaluated separately and in the same manner as the primary model under evaluation. A sub-model contact for additional information should be provided. Sub-model SMEs may need to be brought into the conversation with regulators (whether in-house or third-party sub-models are used).
A.2.d	If using catastrophe model output, identify the vendor and the model settings/assumptions used when the model was run.	1	To accelerate the review of the filing, get contact information for the SME that ran the model and an SME from the vendor. The "SME" can be an intermediary at the insurer (e.g., a filing specialist), who can place the regulator in direct contact with the appropriate SMEs at the insurer or model vendor. For example, it is important to know hurricane model settings for storm surge, demand surge, and long-term/short-term views.
A.2.e	Obtain an explanation of how catastrophe models are integrated into the model to ensure no double-counting.	1	If a weather-based sub-model is input to the Tree-based model under review, loss data used to develop the model should not include loss experience associated with the weather-based sub-model. Doing so could cause distortions in the modeled results by double-counting such losses when determining relativities or loss loads in the filed rating plan. For example, redundant losses in the data may occur when non-hurricane wind losses are included in the data while also using a severe convective storm model in the actuarial indication. Such redundancy may also occur with the inclusion of fluvial or pluvial flood losses when using a flood model or inclusion of freeze losses when using a winter storm model.
A.2.f	If using output of any scoring algorithms, obtain a list of the variables used to determine the score, and provide the source of the data used to calculate the score.	1	Any sub-model should be reviewed in the same manner as the primary model that uses the sub-model's output as input. Depending on the result of item A.2.b, the importance of this item may be decreased.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
3. Adjustments to Data			
A.3.a	Determine if premium, exposure, loss, or expense data were adjusted (e.g., on-leveled, developed, trended, adjusted for catastrophe experience, or capped). If so, how? Do the adjustments vary for different segments of the data? If so, identify the segments and how the data was adjusted.	2	The rating plan or indications underlying the rating plan may provide special treatment of large losses and non-modeled large loss events. If such treatments exist, the company should provide an explanation of how they were handled. These treatments need to be identified, and the company/regulator needs to determine whether model data needs to be adjusted. For example, should large bodily injury (BI) liability losses in the case of personal automobile insurance be excluded, or should large non-catastrophe wind/hail claims in home insurance be excluded from the model's training, test, and validation data? Look for anomalies in the data that should be addressed. For example, is there an extreme loss event in the data? If other processes were used to load rates for specific loss events, how is the impact of those losses considered? Examples of losses that can contribute to anomalies in the data are large losses or flood, hurricane, or severe convective storm losses for personal automobile comprehensive or home insurance. Premium should be brought to current rate level if the target variable is calculated with a premium metric, such as loss ratio. Premium can be brought to current rate level with the extension of exposures method or the parallelogram method. Note that the premium must be on-leveled at a granular variable level for each variable included in the new model if the parallelogram method is used. Statewide on-level factors by coverage are typically sufficient for statewide rate indication development but not sufficient for models that determine rates by variable level.
A.3.b	Identify adjustments that were made to aggregated data (e.g., transformations, binning, and/or categorizations). If any, identify the name of the characteristic/variable, and obtain a description of the adjustment.	1	Pre-modeling binning may be unnecessary in a Tree-based model. The tree model will naturally segment numerical values in the splitting process of the trees. However, if the insurer does bin variables before modeling, the reason should be understood.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.3.c	Ask for aggregated data (one dataset of pre-adjusted/scrubbed data and one dataset of post-adjusted/scrubbed data) that allows the regulator to focus on the univariate distributions and compare raw data to adjusted/binned/transformed/etc. data.	4	This is most relevant for variables that have been "scrubbed" or adjusted. Though most regulators may never ask for aggregated data and do not plan to rebuild any models, a regulator may ask for this aggregated data or subsets of it. It would be useful to the regulator if the percentage of exposures and premium for missing information from the model data by category are provided. This data can be displayed in either graphical or tabular formats.
A.3.d	Determine how missing data was handled.	1	This is most relevant for variables that have been "scrubbed" or adjusted. The regulator should be aware of assumptions the modeler made in handling missing, null, or "not available" values in the data. For example, it would be helpful to the reviewer if the modeler were to provide a statement as to whether there is any systemic reason for missing data. If adjustments or recoding of values were made, they should be explained. It may also be useful to the regulator if the percentage of exposures and premium for missing information from the model data are provided. This data can be displayed in either graphical or tabular formats. The modeler should describe the way the tree fitting process handled missing values. The modeler should specify if missing values are treated before running the Tree-based model or if they are allowed to be handled by the Tree-based model. When creating predictions on new datasets (such as hold out datasets), tree-based models may have different approaches for handling missing data or categorical levels not encountered in the training data for a predictor variable. The modeler should specify the process utilized when this occurs.
A.3.e	If duplicate records exist, determine how they were handled.	1	
A.3.f	Determine if there were any material outliers identified and subsequently adjusted during the scrubbing process.	3	Look for a discussion of how outliers were handled. If necessary, the regulator may want to investigate further by getting a list (with description) of the types of outliers, and determine what adjustments were made to each type of outlier. To understand the filer's response, the regulator should ask for the filer's materiality standard.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
4. Data Organization			
A.4.a	Obtain documentation on the methods used to compile and organize data, including procedures to merge data from different sources or filter data based on particular characteristics and a description of any preliminary analyses, data checks, and logical tests performed on the data and the results of those tests.	2	This should explain how data from separate sources was merged and/or how subsets of policies, based on selected characteristics, are filtered to be included in the data underlying the model and the rationale for that filtering.
A.4.b	Obtain documentation on the insurer's process for reviewing the appropriateness, reasonableness, consistency, and comprehensiveness of the data, including a discussion of the rational relationship the data has to the predicted variable.	2	An example is when by-peril or by-coverage modeling is performed; the documentation should be for each peril/coverage and make rational sense. For example, if "murder" or "theft" data is used to predict the wind peril, the company should provide support and a rational explanation for their use.
A.4.c	Identify material findings the company had during its data review, and obtain an explanation of any potential material limitations, defects, bias, or unresolved concerns found or believed to exist in the data. If issues or limitations in the data influenced modeling analysis and/or results, obtain a description of those concerns and an explanation how modeling analysis was adjusted and/or results were impacted.	1	"None" or "N/A" may be an appropriate response.

B. BUILDING THE MODEL

Section	Information Element	Level of Importance to the Regulator's Review	Comments
1. High-Level Narrative for Building the Model			
B.1.a	Identify the type of model underlying the rate filing (e.g., Random Forest, GLM, decision tree, Bayesian GLM, gradient-boosting machine, neural network, etc.). Understand the model's role in the rating system and provide the reasons why that type of model is an appropriate choice for that role.	1	It is important to understand if the model in question is a Tree-based model and, therefore, these information elements are applicable, or if it is some other model type, in which case other reasonable review approaches may be considered. There should be an explanation of why the model (using the variables included in it) is appropriate for the line of business. If by-peril or by-coverage modeling is used, the explanation should be by- peril/by-coverage. **Note**: If the model is not a Tree-based model, the information elements in this appendix may not apply in their entirety.
B.1.b	Identify the software used for model development. Obtain the name of the software vendor/developer, software product, and a software version reference used in model development.	3	Changes in software from one model version to the next may explain if such changes, over time, contribute to changes in the modeled results. The company should provide the name of the third-party vendor and a contact in the event the regulator has questions. The contact can be an intermediary at the insurer (e.g., a filing specialist) who can place the regulator in direct contact with the appropriate SME at the vendor. Open-source software/programs used in model development should be identified by name and version the same as if from a vendor.
B.1.c	Obtain a description of how the available data was divided between model training, test, and/or validation datasets. The description should include an explanation why the selected approach was deemed most appropriate, whether the company made any further subdivisions of available data, and reasons for the subdivisions (e.g., a portion separated from training data to support testing of components during model building). Determine if the validation data was accessed before model training was completed and, if so, obtain an explanation of why that came to occur. Obtain a discussion of whether the model was rebuilt using all the data or if it was only based on the training data.	1	The reviewer should be aware that modelers may break their data into three or just two datasets. Although the term "training" is used with little ambiguity, "test" and "validation" are terms that are sometimes interchanged, or the word "validation" may not be used at all. The reviewer should note whether a company employed cross-validation techniques instead of a training/test/validation dataset approach. If cross-validation techniques were used, the reviewer should request a description of how cross-validation was done and confirm that the final model was not built on any particular subset of the data, but rather the full dataset. The discussion of training, test, and/or validation datasets is a separate discussion from the percentage of observations (rows of data) or percentage of features (columns of data) used within each tree. These splits are based on hyperparameters and are commented on in other sections.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.1.d	Obtain a brief description of the development process, from initial concept to final model and filed rating plan.	1	The narrative should have the same scope as the filing.
B.1.e	Obtain a narrative on whether loss ratio, pure premium, or frequency/severity analyses were performed and, if separate frequency/severity modeling was performed, how pure premiumswere determined.	1	
B.1.f	Identify the model's target variable.	1	A clear description of the target variable is key to understanding the purpose of the model. It may also prove useful to obtain a sample calculation of the target variable in Excel format, starting with the "raw" data for a policy, or a small sample of policies, depending on the complexity of the target variable calculation.
B.1.g	Obtain a description of the candidate variable selection process prior to the model building.	1	Candidate variables are the variables used as input to the modeling process. Certain variables may not end up used in the final model if none of the component trees of the model split on the variable. The narrative regarding the candidate variable selection process may address matters such as the criteria upon which variables were selected or omitted, identification of the number of preliminary variables considered in developing the model versus the number of variables that remained, and any statutory or regulatory limitations that were taken into account when making the decisions regarding candidate variable selection. The modeler should comment on the use of automated feature selection algorithms to choose candidate predictor variables and explain how potential overfitting that can arise from these techniques was addressed.
B.1.h	In conjunction with variable selection, obtain a narrative on how the company determined the granularity of the rating variables during model development.	3	The narrative should include discussion of how credibility was considered in the process of determining the level of granularity of the variables selected.
B.1.i	Determine if model input data was segmented in any way (e.g., by-coverage, by-peril, or by-form basis). If so, obtain a description of data segmentation and the reasons for data segmentation.	1	The regulator would use this to follow the logic of the modeling process.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
2. Medium-Level Narrative for Building the Model			
B.2.a	At crucial points in model development, if selections were made among alternatives regarding model assumptions, techniques, or hyperparameters, obtain a narrative on the judgment used to make those selections.	2	
B.2.b	If post-model adjustments were made to the data and the model was rerun, obtain an explanation on the details and the rationale for those adjustments.	2	Evaluate the addition or removal of variables and the model fitting. It is not necessary for the company to discuss each iteration of adding and subtracting variables, but the regulator should gain a general understanding of how these adjustments were done, including any statistical improvement measures relied upon.
B.2.c	Identify which distribution was used for the model (e.g., Regression based on Poisson, Gamma, Logistic, or Tweedie are common choices). Obtain an explanation of why the distribution was chosen. Certain distribution assumptions will involve numerical parameters; i.e., regression with a Tweedie assumed distribution will have a p power value. Obtain the specific numerical parameters associated with the distribution.	1	
B.2.d	Obtain a narrative on how the predictions from the component trees are combined to arrive at a final model prediction.	2	Tree-based methods combine predictions from multiple component trees and aggregate them into a final prediction for each observation. Common methods for combining Random Forest model predictions include the arithmetic or geometric mean of all the component trees. Boosting algorithms further refine the model iteratively in each tree, with a focus on records where predictions were off in prior iterations. Gradient Boosting Machines similarly aggregate predictions from all trees. Producing predictions sometimes involve summing all applicable terminal node values and applying the inverse of a link function.
B.2.e	If there were data situations in which weights were used, obtain an explanation of how and why they were used.	3	Investigate whether identical records were combined to build the model.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.2.f	Obtain the number of component trees comprising the Tree-based model. Obtain a narrative on how this number was chosen.	1	Tree-based models should contain enough trees to reduce error to an acceptable level. They should also balance this with the concept of parsimony. A model with fewer trees that achieves relatively similar reduction in error is preferable to a model with more trees. Checking the error on a test dataset or out of bag error for different numbers of trees can reveal at what value the error on test data starts to level off. Modelers might rely on early stopping rules within modeling software to arrive at the final number of trees. The narrative on the number of trees should discuss the stopping criterion, which defines what condition is met when the model stopped adding more trees.
B.2.g	Obtain the sampling parameters that apply to both the percent of observations used in each component tree and the number of features tested for each split within each tree. Obtain a narrative on how the sampling parameters were selected.	1	Tree-based models often sample both the observations (typically rows of modeling data) with replacement and sample the features (typically columns of modeling data) This means that each tree has a bootstrapped dataset. The company should discuss the bagging fraction (sample size) applied to observations (typically rows of data). This is often expressed as a percent. For example: perhaps each tree is based on a bootstrapped sample that is 50% of the original dataset. The company should discuss the number of features considered at each split. This is often expressed as an integer. A common choice for the number of features is equal to roughly the square root of the total number of candidate variables. For example: perhaps each split is based on 10 randomly selected features (typically columns of data) when there are 100 candidate variables.
B.2.h	Obtain the maximum depth that applies to the component trees in the model. Obtain a narrative on how this number was chosen.	1	The depth of a tree is the number of splits that are allowed to occur between the root node and the terminal nodes. This number can be set explicitly in modeling software or may be implicitly set if the company applies a splitting constraint, such as a minimum observations per node. Maximum tree depths of eight or higher are considered extremely high.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.2.i	Obtain parameters that determined the volume of data in each tree node and a narrative of how parameters were chosen.	1	Minimum data volume constraints can be applied to a tree-based model, such that the trees will not create a split that would result in terminal nodes with volume below a set amount. The modeler should comment on how the threshold was chosen. If there was no minimum data volume threshold applied to the trees, or if the threshold was exceedingly small, obtain an explanation of any post-modeling adjustments the modeler made to address the credibility considerations and how the adjustments were applied.
B.2.j	Obtain the learning rate aka "shrinkage" if the model is a Gradient Boosting Machine	1	Learning rate is a hyperparameter that applies to Gradient Boosting Machines but not to random forest models. The hyperparameter controls how far towards indicated each tree is allowed to move. The number is typically set to a low value, to reflect that GBM is intended to be a collection of "weak learners", whose accuracy comes after ensembling a large number of trees. As a rule of thumb, values less than or equal to 0.20 are common.
B.2.k	Obtain a narrative of the process to select all hyperparameters for the Tree-based model. Detail how this process addressed potential overfitting in the model.	2	The narrative should include a description of each hyperparameter, document the values of the hyperparameters, specify the implication of using a higher or lower value for each hyperparameter, and discuss any sensitivity testing completed on the hyperparameters and observations from the sensitivity analysis. Hyperparameter tuning can be done in a variety of ways. The rigor of the tuning process should reflect the risk of overfitting on the specific dataset.

3. Predictor Variables

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.3.a	Obtain a complete data dictionary, including the names, types, definitions, and rationales for each variable.	1	Types of variables might be continuous, discrete, Boolean, etc. Identify any variable used as an offset or control in the Tree-based model and the offset factor that was applied for each level of the offset variable. For any variable(s) intended to function as a control or offset, obtain an explanation of its purpose and impact. Also, for any use of interaction between variables, obtain an explanation of its rationale and impact.
B.3.b	Obtain a list of predictor variables considered but not used in the final model and the rationale for their removal.	4	The purpose of this requirement is to identify variables the company finds to be predictive but ultimately may reject for reasons other than loss-cost considerations (e.g., price optimization). Also, look for variables the company tested and then rejected. This item could help address concerns about data dredging.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.3.c	Obtain a correlation matrix for all predictor variables included in the model and sub-model(s).	3	High correlation is less of an issue for tree-based models than it is for GLMs. Tree-based models naturally only use one variable at a time during each split in each tree. However, a correlation matrix still helps the reviewer understand relationships in the data being modeled better. The company should indicate what statistic was used (e.g., Pearson, Cramer's V, etc.) in the correlation matrix. The regulatory reviewer should understand what statistic was used to produce the matrix but should not prescribe the statistic.
B.3.d	Obtain plots describing the relationship between each predictor variable and the target variable. Obtain a rational explanation for the observed relationship between each predictor variable and the target variable (frequency, severity, loss costs, expenses, or any element or characteristic being predicted).	1	Partial dependence plots (PDPs), accumulated local effects (ALE) plots, or Shapley plots will help improve model interpretability. There should be at least one plot for every variable used in the model. The plots should be accompanied by commentary on why the visualized relationship is reasonable for variables of concern. Considering possible causation may be relevant, but proving causation is neither practical nor expected. If no rational explanation can be provided, greater scrutiny may be appropriate.

For example, the regulator should look for unfamiliar predictor variables and, if found, the regulator should seek to understand the relationship that variable has to the target variable.

The regulator should also consider that interpretability plots for tree-based models need to be reviewed with other considerations in mind. For example, partial dependence calculations assume independence with other variables in the model. |
| B.3.e | If the modeler made use of one or more dimensionality reduction techniques, such as a principal component analysis (PCA), obtain a narrative about that process, an explanation why that technique was chosen, and a description of the step-by-step process used to transform observations (usually correlated) into a set of linearly uncorrelated variables. In each instance, obtain a list of the pre-transformation and post-transformation variable names, as well as an explanation of how the results of the dimensionality reduction technique was used within the model. | 2 | |

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.3.f	Obtain variable importance plots. Obtain a description of how variable importance was calculated.	1	Variable Importance Plots for tree-based methods highlight which variables contributed most to the model. There are multiple ways to calculate variable importance. Variables with the lowest importance measures should be prioritized when identifying variables that may not be contributing significantly to the model. Variables may have a low importance measure due to high correlation with other variables but may still prove useful if they interact with other variables to identify unique subsets of risks. Variables with the highest importance measures should be prioritized when determining which variables have the largest impact on predictions.
4. Adjusting Data, Model Validation, and Goodness-of-Fit Measures			
B.4.a	Obtain a description of the methods used to assess the statistical significance/goodness-of-fit of the model to validation data, such as lift charts and statistical tests. Compare the model's projected results to historical actual results and verify that modeled results are reasonably similar to actual results from validation data.	1	For models that are built using multistate data, validation data for some segments of risk is likely to have low credibility in individual states. Nevertheless, some regulators require model validation on state-only data, especially when analysis using state-only data contradicts the countrywide results. State-only data might be more applicable, but it could also be impacted by low credibility for some segments of risk. **Note**: It may be useful to consider geographic stability measures for territories within the state.
B.4.b	Obtain evidence that the model fits the training data well by variable and for the overall model.	2	The regulator should ask for the company to provide exhibits or plots that show the fitted average makes sense when compared to the observed average for variables of interest. Regulators would ideally review this comparison for every variable, but time constraints may limit the focus to just variables of interest. Variables of interest should include those with a high importance measure (which will have the most material impact on rates), those with a low importance measure (which may not be contributing significantly to the model), variables without an intuitive relationship to loss, or variables that may be proxies for a protected class attribute.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.4.c	Obtain a description how the model was tested for stability over time.	2	Evaluate the build/test/validation datasets for potential time-sensitive model distortions (e.g., a winter storm in year 3 of 5 can distort the model in both the testing and validation datasets). Obsolescence over time is a model risk (e.g., old data for a variable or a variable itself may no longer be relevant). If a model being introduced now is based on losses from years ago, the reviewer should be interested in knowing whether that model would be predictive in the proposed context. Validation using recent data from the proposed context might be requested. Obsolescence is a risk even for a new model based on recent and relevant loss data. The reviewer may want to inquire as to the following: What steps, if any, were taken during modeling to prevent or delay obsolescence? What controls exist to measure the rate of obsolescence? What is the plan and timeline for updating and ultimately replacing the model? The reviewer should also consider that as newer technologies enter the market (e.g., personal automobile), their impact may change claim activity over time (e.g., lower frequency of loss). So, it is not necessarily a bad thing that the results are not stable over time.
B.4.d	Obtain a narrative on how potential concerns with overfitting were addressed.	2	Tree-based models are notorious for overfitting. The company should provide a narrative on how overfitting was addressed. The company should provide a lift chart on training data used to fit the model and a lift chart on testing data that was not used to fit the model. If pruning was used to address overfitting, the narrative should provide commentary on the pruning process.
B.4.e	Obtain support demonstrating that the model assumptions are appropriate.	3	A visual review of plots of actual errors is usually sufficient. The reviewer should look for a conceptual narrative covering these topics: How does this particular Tree-based model work? Why did the rate filer do what they did? Why employ this design instead of alternatives? Why choose this particular distribution function and this particular link function? A company response may be at a fairly high level and reference industry practices. If the reviewer determines that the model makes no assumptions that are considered to be unreasonable, the importance of this item may be reduced.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.4.f	Obtain 5-10 sample records with corresponding output from the model for those records.	2	The company should provide comprehensive documentation of the rating algorithm such that a rate can be reproduced for any theoretical risk. The company should demonstrate the comprehensiveness of the documentation by providing 5-10 sample records with corresponding input variable values and the final model prediction. The company should describe how the final model prediction aggregates the individual tree model predictions. The company should describe how to use other filing exhibits to reproduce the final model prediction for each sample record.
B.4.g	Obtain a deviance analysis by number of trees.	2	The company should provide a plot showing that the deviance of the overall model decreases after each iteration (each additional tree). Plots which show negative log-likelihood would also be sufficient as models which minimize negative log-likelihood also minimize deviance. If the company chooses an error metric other than deviance or log-likelihood, the company should describe why they chose a different metric and explain how it is calculated.
5. "Old Model" Versus "New Model"			
B.5.a	Obtain an explanation of why this model is an improvement to the current rating plan. If it replaces a previous model, find out why it is better than the one it is replacing; determine how the company reached that conclusion and identify metrics relied on in reaching that conclusion. Look for an explanation of any changes in calculations, assumptions, parameters, and data used to build this model from the previous model.	2	The regulator should expect to see improvement in the new class plan's predictive ability or other sufficient reason for the change.
B.5.b	Determine if two Gini coefficients were compared and obtain a narrative on the conclusion drawn from this comparison.	3	This information element requests a comparison of the Lorenz curve and Gini coefficient from the prior model to the Gini coefficient of proposed model. It is expected that there should be improvement in the Gini coefficient. A higher Gini coefficient indicates greater differentiation produced by the model and how well the model fits that data. This is relevant when one model is being updated or replaced. The regulator should expect to see improvement in the new class plan's predictive ability. One example of a comparison might be sufficient. Note: This comparison is not applicable to initial model introduction. The reviewer can look to CAS monograph, "Generalized Linear Models for Insurance Rating."

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.5.c	Determine if double-lift charts were analyzed and obtain a narrative on the conclusion drawn from this analysis.	3	One example of a comparison might be sufficient. Note: "Not applicable" is an acceptable response.
B.5.d	If replacing an existing model, obtain a list of any predictor variables used in the old model that are not used in the new model as candidate variables. Obtain an explanation of why these variables were dropped from the new model. Obtain a list of all new predictor variables in the new model that were not in the prior old model.	2	It is useful to differentiate between old and new variables so the regulator can prioritize more time on variables not yet reviewed.
6. Modeler Software			
B.6.a	Request access to SMEs (e.g., modelers) who led the project, compiled the data, and/or built the model.	4	The filing should contain a contact that can put the regulator in touch with appropriate SMEs and key contributors to the model development to discuss the model.

C. THE FILED RATING PLAN

Section	Information Element	Level of Importance to the Regulator's Review	Comments
1. General Impact of Model on Rating Algorithm			
C.1.a	In the actuarial memorandum or explanatory memorandum, for each model and sub-model (including external models), look for a narrative that explains each model and its role (i.e., how it was used) in the rating system.	1	The "role of the model" relates to how the model integrates into the rating plan as a whole and where the effects of the model are manifested within the various components of the rating plan. This is not intended as an overarching statement of the model's goal, but rather a description of how specifically the model is used. This item is particularly important if the role of the model cannot be immediately discerned by the reviewer from a quick review of the rate and/or rule pages. (Importance is dependent on state requirements and ease of identification by the first layer of review and escalation to the appropriate review staff.)
C.1.b	Obtain an explanation of how the model was used to adjust the filed rating algorithm.	1	The regulator should consider asking for an explanation of how the model was used to adjust the rating algorithm.
C.1.c	Obtain a complete list of characteristics/variables used in the proposed rating plan, including those used as input to the model (including sub-models and composite variables) and all other characteristics/ variables (not input to the model) used to calculate a premium. For each characteristic/variable, determine if it is only input to the model, whether it is only a separate univariate rating characteristic, or whether it is both input to the model and a separate univariate rating characteristic. The list should include transparent descriptions (in plain language) of each listed characteristic/variable.	1	Examples of variables used as inputs to the model and used as separate univariate rating characteristics might be criteria used to determine a rating tier or household composite characteristic.
2. Relevance of Variables and Relationship to Risk of Loss			
C.2.a	Obtain a narrative regarding how the characteristics/rating variables included in the filed rating plan relate to the risk of insurance loss (or expense) for the type of insurance product being priced.	2	The narrative should include a discussion of the relevance each characteristic/rating variable has on consumer behavior that would lead to a difference in risk of loss (or expense). The narrative should include a rational relationship to cost, and model visualization plots (such as partial dependence plots, accumulated local effects plots, or Shapley plots) should be consistent with the expected direction of the relationship. **Note**: This explanation would not be needed if the connection between variables and risk of loss (or expense) has already been illustrated.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
3. Comparison of Model Outputs to Current and Selected Rating Factors			
C.3.a	Obtain documentation and support for all calculations, judgments, or adjustments that connect the model's indicated values to the selected rates filed in the rating plan.	1	The documentation should include explanations for the necessity of any such adjustments and each significant difference between the model's indicated values and the selected values. This applies even to models that produce scores, tiers, or ranges of values for which indications can be derived. **Note**: This information is especially important if differences between model-indicated values and selected values are material and/or impact one consumer population more than another.
C.3.b	For each characteristic/variable used as both input to the model (including sub-models and composite variables) and as a separate univariate rating characteristic, obtain a narrative regarding how each characteristic/variable was tempered or adjusted to account for possible overlap or redundancy in what the characteristic/variable measures.	2	The insurer should address this possibility or other considerations; e.g., tier placement models often use risk characteristics/variables that are also used elsewhere in the rating plan. One way to do this would be to model the loss ratios resulting from a process that already uses univariate rating variables. Then the model/composite variables would be attempting to explain the residuals.
4. Responses to Data, Credibility, and Granularity Issues			
C.4.a	Determine what, if any, consideration was given to the credibility of the output data.	2	The regulator should determine at what level of granularity credibility is applied. If modeling was by coverage, by form, or by peril, the company should explain how these were handled when there was not enough credible data by coverage, form, or peril to model.
C.4.b	If the rating plan is less granular than the model, obtain an explanation of why.	2	This is applicable if the company had to combine modeled output in order to reduce the granularity of the rating plan.
C.4.c	If the rating plan is more granular than the model, obtain an explanation of why.	2	A more granular rating plan may imply that the company had to extrapolate certain rating treatments, especially at the tails of a distribution of attributes, in a manner not specified by the model indications. It may be necessary to extrapolate due to data availability or other considerations.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
5. Definitions of Rating Variables			
C.5.a	Obtain a narrative regarding adjustments made to model output (e.g., transformations, binning and/or categorizations). If adjustments were made, obtain the name of the characteristic/variable and a description of the adjustment.	2	If rating tiers or other intermediate rating categories are created from model output, the rate and/or rule pages should present these rating tiers or categories. The company should provide an explanation of how model output was translated into these rating tiers or intermediate rating categories.
6. Supporting Data			
C.6.a	Obtain aggregated state-specific, book-of-business-specific univariate historical experience data, separately for each year included in the model, consisting of loss ratio or pure premium relativities and the data underlying those calculations for each category of model output(s) proposed to be used within the rating plan. For each data element, obtain an explanation of whether it is raw or adjusted and, if the latter, obtain a detailed explanation for the adjustments.	4	For example, were losses developed/undeveloped, trended/untrended, capped/uncapped, etc.? Univariate indications should not necessarily be used to override more sophisticated multivariate indications. However, they do provide additional context and may serve as a useful reference.
7. Consumer Impacts			
C.7.a	Obtain a listing of the top five rating variables that contribute the most to large swings in renewal premium, both as increases and decreases, as well as the top five rating variables with the largest spread of impact for both new and renewal business.	4	These rating variables may represent changes to rating factors, be newly introduced to the rating plan, or have been removed from the rating plan.
C.7.b	Determine if the company performed sensitivity testing to identify significant changes in premium due to small or incremental change in a single risk characteristic. If such testing was performed, obtain a narrative that discusses the testing and provides the results of that testing.	3	One way to see sensitivity is to analyze a graph of each risk characteristic's/variable's average fitted model prediction. Look for significant variation between the average fitted model predictions for adjacent rating variable levels and evaluate if such variation is reasonable and credible.
C.7.c	For the proposed filing, obtain the impacts on renewal business, and describe the process used by management, if any, to mitigate those impacts.	2	Some mitigation efforts may substantially weaken the connection between premium and expected loss and expense and, hence, may be viewed as unfairly discriminatory by some states.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.7.d	Obtain a rate disruption/dislocation analysis demonstrating the distribution of percentage and/or dollar impacts on renewal business (created by rerating the current book of business) and sufficient information to explain the disruptions to individual consumers.	2	The analysis should include the largest dollar and percentage impacts arising from the filing, including the impacts arising specifically from the adoption of the model or changes to the model as they translate into the proposed rating plan. While the default request would typically be for the distribution/dislocation of impacts at the overall filing level, the regulator may need to delve into the more granular variable-specific effects of rate changes if there is concern about particular variables having extreme or disproportionate impacts, or significant impacts that have otherwise yet to be substantiated. See Appendix D for an example of a disruption analysis.
C.7.e	Obtain exposure distributions for the model's output variables and show the effects of rate changes at granular and summary levels, including the overall impact on the book of business.	3	See Appendix D for an example of an exposure distribution.
C.7.f	Identify policy characteristics, used as input to a model or sub-model, that remain "static" over a policy's lifetime versus those that will be updated periodically. Obtain a narrative on how the company handles policy characteristics that are listed as "static," yet change over time.	3	Some examples of "static" policy characteristics are prior carrier tenure, prior carrier type, prior liability limits, claim history over past X years, or lapse of coverage. These are specific policy characteristics usually set at the time new business is written, used to create an insurance score or to place the business in a rating/underwriting tier, and often fixed for the life of the policy. The reviewer should be aware of, and possibly concerned about, how the company treats an insured over time when the insured's risk profile based on "static" variables changes over time, but the rate charged, based on a new business insurance score or tier assignment, no longer reflect the insured's true and current risk profile. A few examples of "non-static" policy characteristics are age of driver, driving record, and credit information (FCRA-related). These are updated automatically by the company on a periodic basis, usually at renewal, with or without the policyholder explicitly informing the company.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.7.g	Obtain a means to calculate the rate charged a consumer.	3	The filed rating plan should contain enough information for a regulator to be able to validate policy premium. However, for a complex model or rating plan, a score or premium calculator via Excel or similar means would be ideal, but this could be elicited on a case-by-case basis. The ability to calculate the rate charged could allow the regulator to perform sensitivity testing when there are small changes to a risk characteristic/variable. Note: This information may be proprietary. For the rating plan, the rate order of calculation rule may be sufficient. However, it may not be feasible for a regulator to get all the input data necessary to reproduce a model's output. Credit and telematics models are examples of model types where model output would be readily available, but the input would not be readily available to the regulator.
C.7.h	In the filed rating plan, be aware of any non-insurance data used as input to the model (customer-provided or other). In order to respond to consumer inquiries, it may be necessary to inquire as to how consumers can verify their data and correct errors.	1	If the data is from a third-party source, the company should provide information on the source. Depending on the nature of the data, it may need to be documented with an overview of who owns it. The topic of consumer verification may also need to be addressed, including how consumers can verify their data and correct errors.
8. Accurate Translation of Model into a Rating Plan			
C.8.a	Obtain sufficient information to understand how the model outputs are used within the rating system and to verify that the rating plan's manual, in fact, reflects the model output and any adjustments made to the model output.	1	The regulator can review the rating plan's manual to see that modeled output is properly reflected in the manual's rules, rates, factors, etc.
9. Efficient and Effective Review of Rate Filing			
C.9.a	Establish procedures to efficiently review rate filings and models contained therein.	1	"Speed to market" is an important competitive concept for insurers. Although the regulator needs to understand the rate filing before accepting the rate filing, the regulator should not request information that does not increase his/her understanding of the rate filing. The regulator should review the state's rate filing review process and procedures to ensure that they are fair and efficient.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.9.b	Be knowledgeable of state laws and regulations in order to determine if the proposed rating plan (and models) are compliant with state laws and/or regulations.	1	This is a primary duty of state insurance regulators. The regulator should be knowledgeable of state laws and regulations and apply them to a rate filing fairly and efficiently. The regulator should pay special attention to prohibitions of unfair discrimination.
C.9.c	Be knowledgeable of state laws and regulations in order to determine if any information contained in the rate filing (and models) should be treated as confidential.	1	The regulator should be knowledgeable of state laws and regulations regarding confidentiality of rate filing information and apply them to a rate filing fairly and efficiently. Confidentiality of proprietary information is key to innovation and competitive markets.
C.9.d	Obtain complete documentation that would allow future audits of model predictions.	1	The company should provide comprehensive documentation of the rating algorithm such that a rate can be reproduced for any theoretical risk. Comprehensive documentation could be provided as one of the following: a complete set of tree diagrams, a set of if-else logic statements that represents the trees, or a table showing every possible combination of risk characteristics and the final prediction.

TREE-BASED MODELS GLOSSARY OF TERMS

Accumulated Local Effects Plots: A type of interpretability plot. Accumulated local effects (ALE) plots calculate smaller, incremental changes in the feature effects. ALE shows the expected and centered effects of a variable.

Bagged Trees: An ensemble of trees where each tree is based on a "bootstrap aggregated" sample.

Branch: A connection on a decision tree between a parent node and a child node. A relationship based on a predictor variable is checked at each node, determining which branch applies.

Candidate Variables: The variables specified by the modeler to be used within the full model. The variable selection process performed by a Tree-based model means that component trees might only use a subset of these variables in each tree.

Child Node: The node below a parent node. The child node is the result of a split that occurs based on a predictor variable. The node above the child node, which is where the split occurred resulting in the creation of the child nodes, is called the parent note. There is one parent node for every child node. The root node is the only node that is not a child node.

Component Tree: An individual tree within an approach based on an ensemble of trees, such as Random Forest or gradient boosting machine.

Deviance: A measure of model fit. Deviance is based on the difference between the log-likelihood of the saturated model and the log-likelihood of the proposed model being evaluated. Smaller values of deviance demonstrate that a model's predictions fit closer to actual. Deviance on training data will always decrease as model complexity increases.

Gradient Boosting Machine: An ensemble of trees model made up a series of "weak learner" trees which iteratively focus more on the residuals of the model at each iterative tree.

Hyperparameter: A model hyperparameter is a model setting specified by the modeler that is external to the model and whose value cannot be estimated from data.

Node: A point on a decision tree. Nodes are either root nodes (the top node), leaf nodes (a terminal node at which point no further splitting occurs), or an internal node that appears in the middle of the tree while splitting is still taking place.

Out-of-Bag Error: Error calculated for observations based on the trees that did not include them in the set of training observations. Out-of-Bag Error is calculable when bootstrapping is used to generate different datasets for each component tree in an ensemble tree method.

Parent Node: The node above a child node. The parent node is where a split occurs based on a predictor variable. The nodes below the parent node, which are a direct result of the parent node's split, are called child nodes. There are typically two child nodes for every parent node. Terminal nodes cannot be parent nodes.

Partial Dependence Plots: A type of interpretability plot. The partial dependence plot computes the marginal effect of a given variable on the prediction.

Pruning: The process of scaling back a tree to reduce its complexity. This results in trees with fewer branches and terminal nodes appearing higher on the tree. Pruning is more common on models built on a single decision tree rather than on ensemble models such as Random Forests or gradient boosting machines.

Random Forest: An ensemble of trees where each tree is based on a bootstrap aggregated sample, and each split is based on a random sample of the candidate variables.

Root Node: The first (top) node in a decision tree. This node contains the entire set of data used by the tree as no splits have occurred yet.

Shapley Additive Explanation Plots: A type of interpretability plot. Shapley plots investigate the effect of including a variable in the model by the order in which it is added. The Shapley value represents the amount the variable of interest contributes to the prediction.

Splitting: The process of dividing a node into two or more sub-nodes, starting from the root node. Splitting occurs at every node up until the terminal (leaf) nodes when the stopping criterion is met.

Stopping Criterion: A criterion applied to the splitting process that informs the node when it is ineligible to split any further. Volume of data is often used as a stopping criterion, such that each leaf node is based on at least a pre-determined amount of data.

Terminal Node: An end node containing no child nodes because the node has met the stopping criterion. The terminal node is associated with a prediction for one of the component trees. The terminal node is also known as a "leaf" node, the resulting endpoint of a decision tree.

Tree-based Model: A model that can be represented as a decision tree or a collection of decision trees.

Tree Depth: The maximum number of splits between the root node and a leaf node for a tree.

Variable Importance: A measure of how the variables (a.k.a. features) contribute to the overall model. There are multiple ways to measure variable importance.