**New Glossary Terms:**

**Accumulated Local Effects Plots:** A type of interpretability plot. Accumulated Local Effects 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 model 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 random variable selection by a random forest 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 1 parent node for every child node. The root node is the only node which is not a child node.

**Component Tree:** An individual tree within an ensemble of trees based method 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.

**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 a internal node which 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 2 child nodes for every parent node. Terminal nodes can not 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 it’s 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 model 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:** Models 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 (aka features) contribute to the overall model. There are multiple ways to measure variable importance.