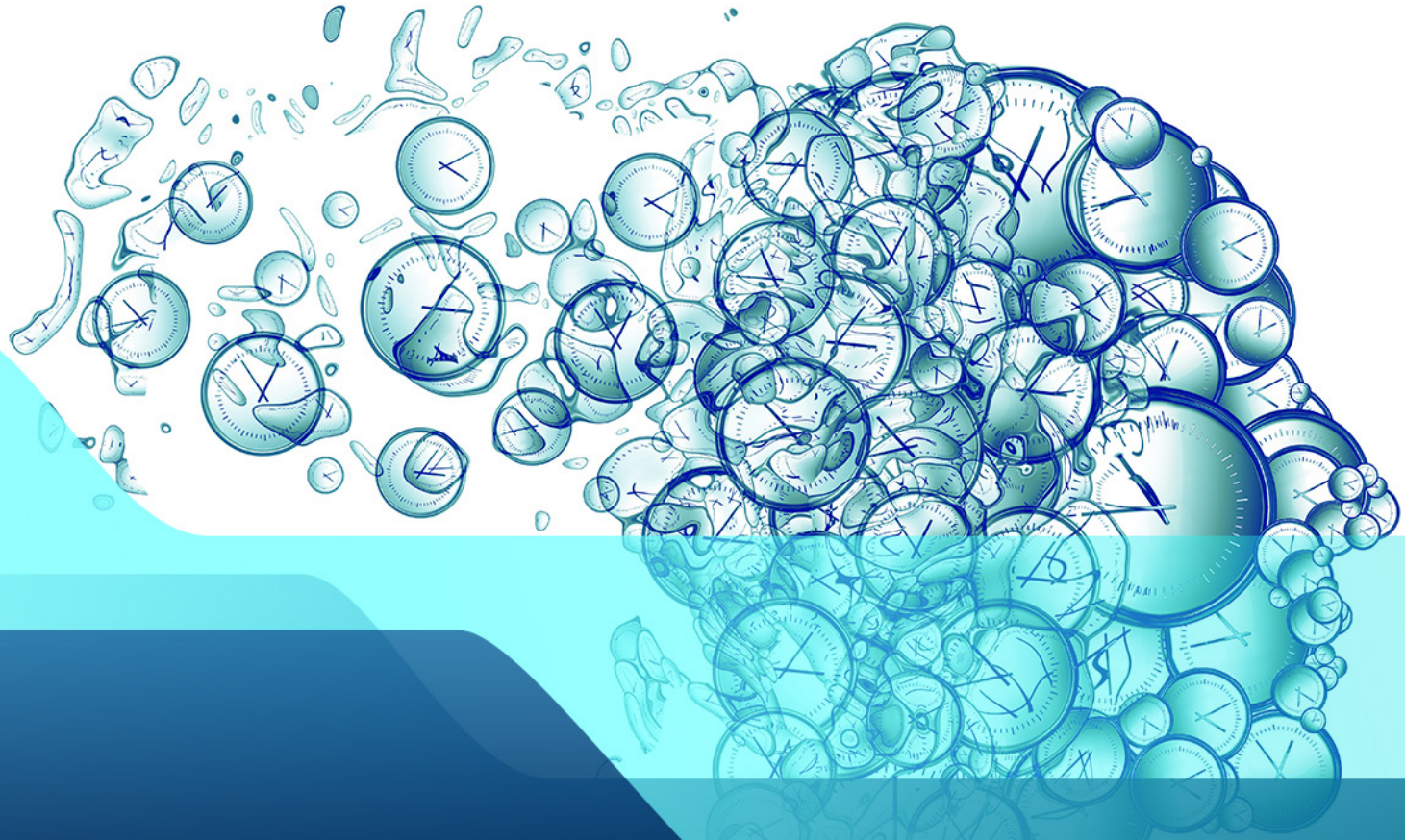
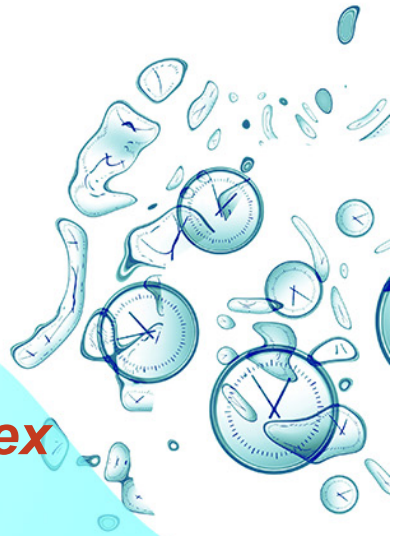


Decision Trees: CART

What is Decision Trees (DT)

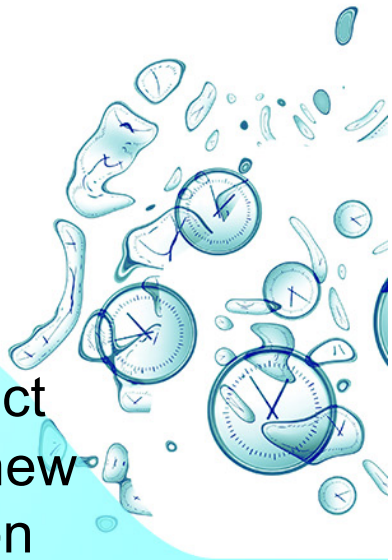
- A Decision Tree is a flow chart of decisions and its outcomes.
- A predictive model based on a branching series of Boolean Tests
- Each internal node tests an attribute
- Each branch corresponds to attribute value
- Each leaf node assigns a classification
- The end nodes can have a category (classification) or a continuous number (regression)
- DT are quite simple and powerful

Algorithms for Decision Trees



- Classification and Regression Trees (CART) uses ***Gini index (classification)*** as metric
- ID3 (Iterative Dichotomiser 3) uses ***Entropy function*** and information gain as metric
- CHAID (Chi-squared Automatic Interaction Detector)

Gini Impurity



- Used by the CART
- Gini Impurity is a measurement of the likelihood of an incorrect classification of a new instance of a random variable, if that new instance were randomly classified according to the distribution of class labels from the data set.
- Consider a dataset D that contains samples from k classes. The probability of samples belonging to class i at a given node can be denoted as p_i. Then the Gini Impurity of D is defined as:

$$Gini(D) = 1 - \sum_{i=1}^{k} p_i^2$$

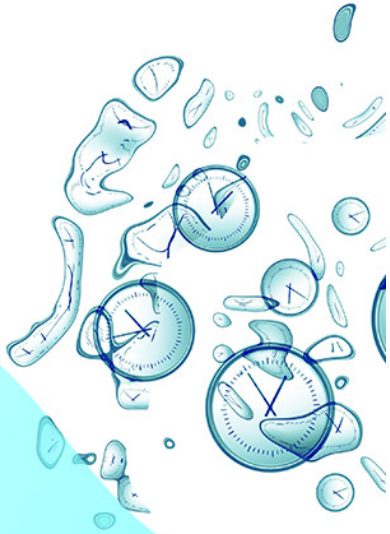
Decision Trees are prone to 'overfitting'

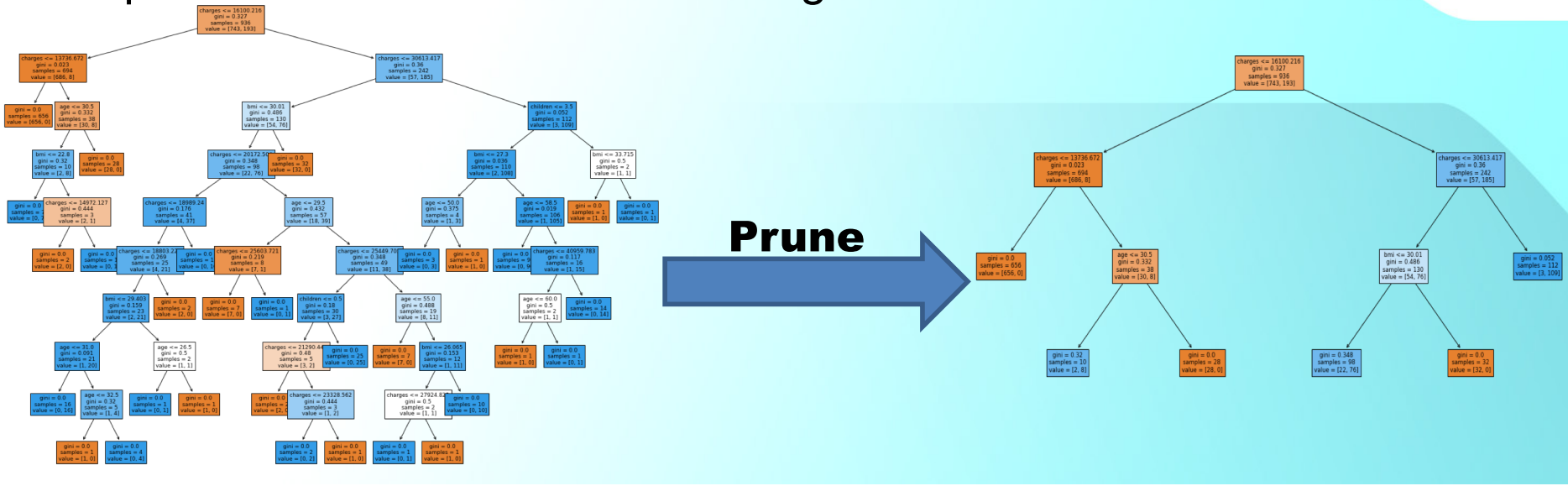
- DT especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions.

- There are several steps to avoiding overfitting in building decision Trees:
  - Pre-pruning that stop growing the tree earlier, before it perfectly classifies the training set.
  - Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

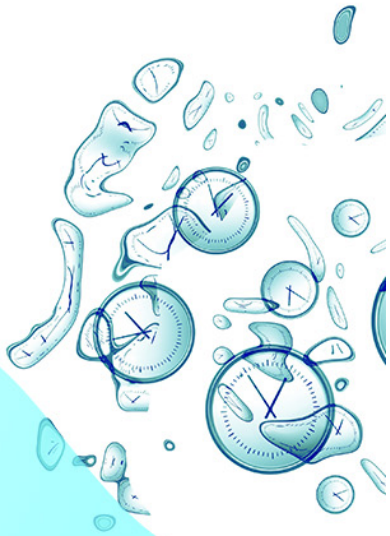
Pruning

- ✓ Pruning is a technique associate with decision trees.
- ✓ Pruning reduces the size of decision trees by removing parts of the tree that do not provide power to classify instances.
- ✓ Simple way to prune a decision tree is by limiting the depth of the tree to avoid over fitting.





Pre-Pruning



- Pre-Pruning prevent the generation of non-significant branches. It stop growing the tree before it grows too big.

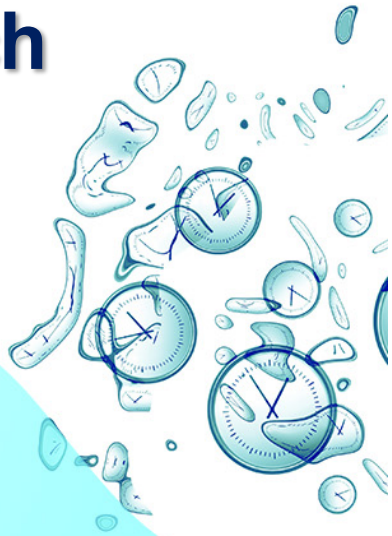
- This technique is used before construction of decision tree

- This can be achieved by bounding hyperparameters:
  - Max depth: The maximum depth of the tree
  - Min samples split: The minimum number of samples required to split a node.
  - Min samples leaf: The minimum number samples required at a leaf node.

Post-Pruning

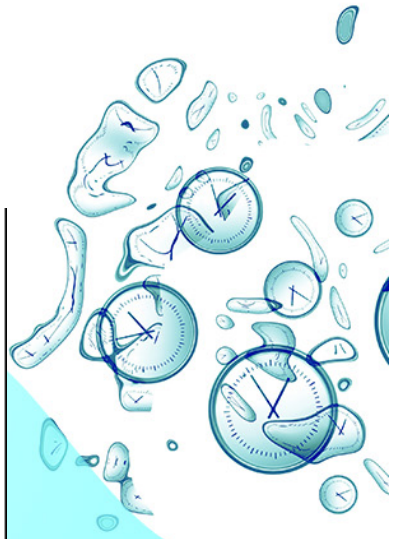
- This technique is used after construction of decision tree
- This technique is used when decision tree will have very large depth and will show overfitting of model.
- It is also known as backward pruning
- This technique is used when we have infinitely grown decision tree.
- At each step the algorithm:
    i. Try to removing each possible subtree
    ii. Find the relative error decrease per node for that subtree-complexity parameter
    iii. And remove the subtree with the minimum

Hyperparameter tuning using Grid Search



a) Grid Search is a process of searching the best combination of hyperparameters from a predefined set of values
b) A parameter grid (Hyperparameters and corresponding values) is provided as an input to the Grid-search function
c) It tries all the combinations of the values passed and evaluates the model for each combination
d) It returns the combination of hyperparameter values that works best as per the metric provided for model evaluation
e) GridSearchCV( ) function is an implementation of Grid Search with Cross Validation

Impurity Measures in Decision Trees



	GINI INDEX	ENTROPY	INFORMATION GAIN	VARIANCE
When to use	Classification	Classification	Classification	Regression
Formula	$1 - \Sigma p_i^2$	$-\Sigma p_i \log(p_i)$	$E(Y) - E(Y\|X)$	$\Sigma(x - \bar{x})^2/N$
Range	0 to 0.5 0 = most pure 0.5 = most impure	0 to 1 0 = most pure 1 = most impure	0 to 1 0 = less gain 1 = more gain	>=0
Characteristics	Easy to compute Non-additive	Computationally intensive Additive	Computationally intensive	The most common measure of spread


Thank You