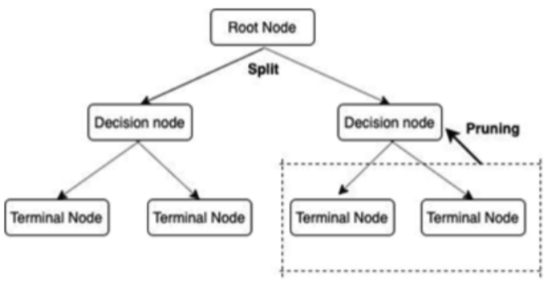
BASIC Terminologies

Description	What	Limitations/Assumptions	Example/Formula	Reference/Comments
Decision Trees	are tree-based models that help in making a decision in both regression and classification problems. Decision trees are considered greedy algorithms	To make a decision, they use a hierarchial structure and split the dataset into smaller subsets		
Entropy	Measure of randomness or impurity contained in a dataset Level of disorder or uncertainity in a given dataset or system High value of entropy means that the randomness in your system is high meaning it is diffcult to predict the state of atoms or molecules in it. Lower value of entropy means that predicting that state is much easier Thus, a node with more variable composition, such as 2Pass and 2 Fail would be considered to have higher Entropy than a node which has only pass or only fail.			
Information gain	Measure of information gained by adding a feature/independent variable or, in other words, reduction in the impurity after adding a feature.			
Root Node	is from where the decision tree starts. It represents the entire population or samples which get divided into two or more branches			
Branch or Sub-Tree	A part of the entire decision tree is called a branch or sub-tree			
Splitting	Dividing a node into two or more sub-nodes based on if-else conditions			
Decision Node	A sub-node that splits into further sub-nodes. In simple terms, every node is a decision node, except for leaf nodes			
Leaf or Terminal Node	End of decision tree where it cannot be split into further sub-nodes			
Depth of the tree	Depth of a decision tree is the number of nodes from the root node to the furthest leaf node			
Greedy Algorithms	is a CS term for any algorithm that ties to approximate the globally optimal solution to a problem by finding the locally optimal solution at each step of the problem instead			

References

https://www.analyticsvidhya.com/blog/2020/11/entropy-a-key-concept-for-all-data-science-beginners/#:~:text=In%20Machine%20Learning%2C%20entropy%20measures,ability%20to%20make%20accurate%20predictions.

https://addepto.com/blog/what-is-entropy-in-machine-learning/#:~:text=In%20other%20words%2C%20a%20high,that%20state%20is%20much%20easier.

https://towardsdatascience.com/decision-trees-explained-entropy-information-gain-gini-index-ccp-pruning-4d78070db36c#:~:text=In%20the%20context%20of%20Decision,only%20pass%20or%20only%20fail.

BASIC Terminologies

Description	What	Limitations/Assumptions	Example/Formula	Reference/Comments

LVC 1 - Glossary of Notations

$\mathcal{X}$ = A vector of categorical data

y = Outcome class (categorical)

$f: \mathcal{X} \rightarrow \mathrm{y}$ = Decision Rule, i.e., f is a function that is mapping the independent features to the target values

x_i = i^{th} row of the vector $\mathcal{X}$

y_i = i^{th} row of vector y

N = Natural number

$\in$ = Belongs to

Σ = The summation

$\neq$ = Not equal to

$R(f)$ = Empirical Error (generalization error) of a Decision Rule

$R^*(f)$ = Probabilistic Error of a Decision Rule

$\frac{1}{N}\sum_{i}^{N} I(f(x_i) \neq y_i)$ = The average number of misclassifications. The $I()$ function is 1 in case of a misclassification and 0 otherwise

C = It is a subclass of data points

case of a misclassification and 0 otherwise

C = It is a subclass of data points

k = Subset of all feature indices in the subclass

Z = Random Variable

X, Y = X represent the independent features and Y represents the target feature

$P(Z)$ = Probability mass function of the random variable Z

E = Expected value

$P(x, y)$ = It represents the joint distribution of X and Y

$H(Z)$ = Entropy of Z

$H(X, Y)$ = Joint Entropy of random variables X and Y

$H(X \mid Y)$ = Conditional Entropy of X given Y

$IG(Y \mid X)$ = Information Gain of Y given X

$X \perp Y$ = X is perpendicular to Y

$X(m)$ = A feature from the X

S_1 = $\{(y_i \mid x_i\,(m) = 0\}$ = Splitting outcome based on class 0

S_2 = $\{(y_i \mid x_i\,(m) = 1\}$ = Splitting outcome based on class 1

BASIC Terminologies				
Description	**What**	**Limitations/Assumptions**	**Example/Formula**	**Reference/Comments**
Decision Tree (DT)	is a supervised learning algorithm used for **classifcation** (spam or not spam) **as well as regression** (pricing a car or a house) problems			
Decision Tree (DT)	is like a **flow chart** where **each internal node represents a test** on an attribute and each branch represents the outcome of that test. In a classification problem, **each leaf node** represents a class label i.e the decision was taken after computing all attributes and the path from the first node to a leaf represents **classification rules** also called **decision rules**			
Advantages of Decision Trees	1. Human-Algorithm Interaction 2. Versatile 3. Built-in Feature selection 4. Testable			
Advantage - Human Algorithm Interaction	1. Simple to understand interpret 2. Mirrors human decision making more closely 3. Uses an open-box model i.e can visualize and understand the machine learning logic (as opposed to a black box model which is not interpretable)			
Advantage- Versatile	1. Able to handle both numerical and categorical data 2. Powerful - can model arbitary functions as long as we have sufficient data 3. Requires little data preparation 4. Performs well with large datasets			
Advantage - Built in feature selection	1. Naturally de-emphasizes irrelevant features 2. Develops a hierarchy in terms of the relevance of features			
Advantage - Testable	Possible to validate a model using statistical tests			
Limitations of Decision Trees	1. Trees can be non-robust 2. Problem of learning an optimal decision tree is known to be NP-Complete 3. Overfitting			
Limitations - Trees can be non-robust	A small change in the training data can result in a large change in the tree and consequently the final predictions			
Limitations - NP-Complete	1. Practical decision tree learning algorithms are based on heuristics (greedy algorithms) 2. Such algorithms cannot guarantee obtaining the globally optimal decision tree			
Limitations - Overfitting	Decision-tree solvers can create over-complex trees that do not generalize well from the training data			
Steps to build a decision tree	Algorthim follows the below steps to build a decision tree 1. Pick a feature 2. Split the data based on that feature that the outcome is binary i.e no data point belongs to both sides of the split 3. Define the new decision rule 4. Repeat the process until each leaf node is homogeneous ie all the data points in a leaf node belong to the same class			