***Decision Tree:***

A decision tree is a predictive modeling tool used in machine learning and data mining. It represents a flowchart-like structure where each internal node represents a decision based on an attribute, each branch represents an outcome of that decision, and each leaf node represents a final decision or prediction. Decision trees are versatile, interpretable, and can handle both classification and regression tasks.

Basic Terminology of Decision Tree:

* Gini Index
* Entropy
* Information Gain

***Gini Index:***

The Gini index, also known as the Gini coefficient or Gini ratio, is a statistical measure of inequality within a distribution. It is commonly

used to quantify the extent of income inequality within a population, but it can also be applied to measure inequality in other contexts such as wealth distribution, resource allocation, or even in evaluating classification models in machine learning.

The Gini index ranges from 0 to 1, where:

A Gini index of 0 represents perfect equality, meaning that every individual in the population has the same income, wealth, or attribute being measured.

A Gini index of 1 represents perfect inequality, where one individual possesses all the income, wealth, or attributes, while everyone else has none.

***How to Gini Index Work.***

The process of utilizing the Gini Index typically involves working with data to construct decision trees. Initially, we compute the Gini index on the dataset, which serves as a pivotal step in the decision tree construction. This index acts as a crucial criterion for determining the optimal splits within the tree nodes, guiding the decision-making process.

Mathematically, it can be expressed as:

G = A / A+B

* *G* is the Gini index.
* *A* is the area between the Lorenz curve and the line of perfect equality.
* *B* is the total area below the line of perfect equality.

| Weekend | Weather\_Session | Parents | Money | Decision |

| W1 | Summy | Yes | Rich | Cinema |

| W2 | Summy | No | Rich | Termins |

| W3 | Windy | Yes | Rich | Cinema |

| W4 | Rainy | No | Poor | Shopping |

| W5 | Windy | Yes | Poor | Staying |

There 4 possible outcome variables   
Cinema, Termins , Shopping , Staying

*Gini*(*Decision*)= 1 - [(⅖)^2 + (⅕)^2 + (⅕)^2 + (⅕) ^ 2 ]

=0.72

***Entropy:***

Entropy, in the context of decision trees, is a measure of the level of disorder or uncertainty within a dataset. It quantifies the impurity of a dataset by calculating the degree of randomness or unpredictability in the distribution of class labels. Higher entropy values indicate greater disorder, while lower values indicate more homogeneity within the dataset.

***Information Gani:***

Information Gain is a concept used in decision tree algorithms to measure the effectiveness of a particular attribute in classifying the data. It represents the reduction in entropy or Gini impurity achieved by splitting the data on a particular attribute. Higher Information Gain suggests that splitting on that attribute results in more homogenous subsets, improving the purity of the resulting nodes in the decision tree.