**DECISION TREE ALGORITHM**

Decision tree is a supervised learning algorithm that is commonly used for solving both classification ad regression but usually preferred for classification problems more. As the name suggests, this model uses tree model (Classification and Regression Tree Model) for representation for labels and attributes. The class labels or outcomes are corresponded by each leaf node, each link symbolises a decision and the inner nodes of the tree represent the attributes or features. This model basically works like a tree node asking question and splitting it into sub-tree and sub-nodes based on the answer.

Although being simple and useful, this model poses a challenge to identify the attribute or feature for root node or sub-node of each level. To solve this problem, Attribute Selection Measure is used which provides us with two popular techniques, **Information Gain** and **Gini Index.** A basic representation of decision tree model can be shown as:



*Reference: Decision Tree by Wikipedia*

Information Gain calculates information provided by a feature about the class and based on that the nodes are further split. It uses uncertainty of a random variable (**entropy)** for the same. Considering S as total number of instances and P as probability, IG can be calculated as:

*Information Gain= Entropy(S) - [(Weighted Average) \* Entropy(each feature)]*

*Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)*

Gini Index measures how often a randomly selected element would turn out incorrect. Lower the GI, more preferred the attribute. GI can be calculated as:

*Gini Index= 1- ∑jPj2*

Common applications of Decision Tree Algorithm include data mining, sentiment analysis, churn analysis and making best decisions such as flights to travel with, restaurants to visit, hotels to stay in.

**Accuracies of DTA of the mentioned dataset:**

1. Voice Gender Recognition –

2. BVC Voice Gender Dataset –