Knowledge Acquisition

It is the process of gathering, selecting and interpreting information to maintain knowledge within a specific domain.

There are different methods of knowledge ~~aq~~ acquisition, including rule based systems, decision trees, artificial neural networks and fuzzy logic systems.

Rule based systems are the simplest form of knowledge based system. They use a set of rules to make decisions.

The most important part of knowledge acquisition is the interpretation of information.

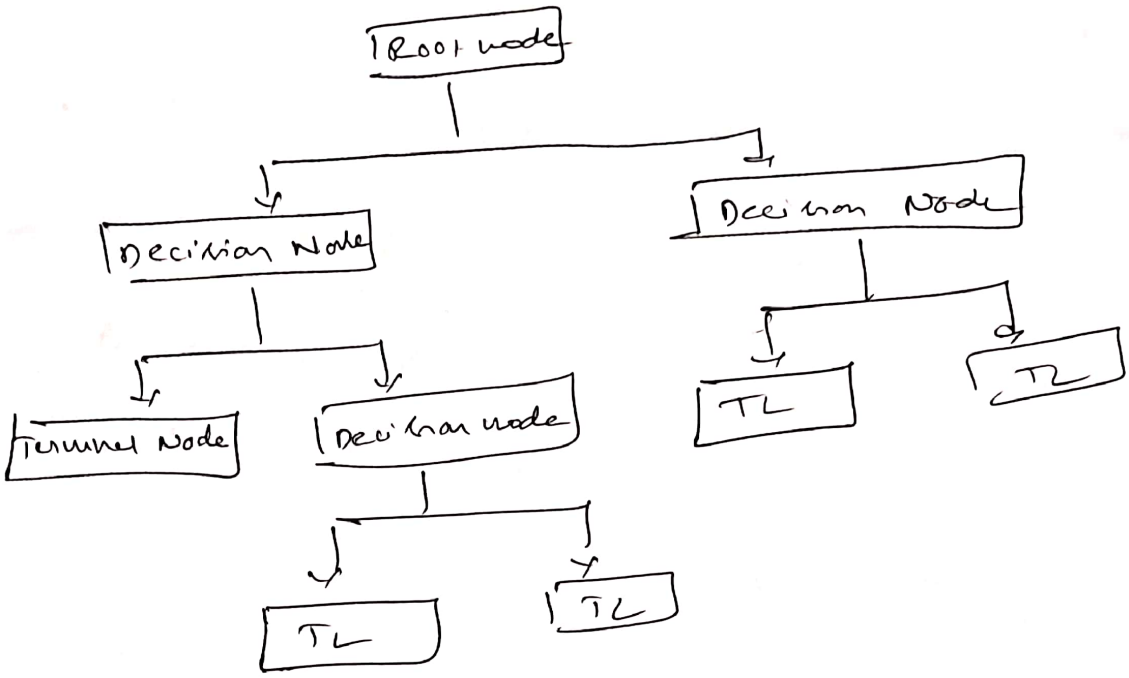
ANN are more complex form of knowledge based systems.

few methods of knowledge Acquisition

1) Expert systems.

2) Learning from examples :- This is a common method used in ML where a system is presented with a set of training data.

Decision Tree :- A Decision Tree is a supervised ML algorithms used for both regression and classification



Root Node :- It is the top most node in tree, which represents the complete dataset.

Decision Node :- Decision nodes are nothing but the result in the spliting of data into multiple data segments.

Leaf node :- Nodes where further Splitting is not possible, often indicating the final classification or outcome

Entropy :- It is used to measure the uncertainty in the data.

ID3 :- Stands for Iterative Dichotomizer 3 features into (repeatedly) (divides) two & node groups.

Splitting Criterias in Decision trees

1) Entropy 2) Gini Index

ID3 Algo :-

	a_1	a_2	a_3	classific
1	True	Hot	High	No
2	T	H	H	N
3	False	H	H	yes
4	F	Cool	Normal	y
5	F	C	N	y
6	T	C	H	N
7	T	H	H	N
8	T	H	N	y
9	F	C	N	y
10	F	C	H	y

Entropy of the entire table

$$S = -\left(\frac{p}{p+n}\right)\log\left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right)\log\left(\frac{n}{p+n}\right)$$

I_G : Entropy - $I(A)$

Entropy of entire table $\{6+, 4-\}$

$$= -\frac{6}{10}\log\frac{6}{10} - \frac{4}{10}\log\frac{4}{10} = 0.9709$$

First attribute

$a_1 = \{T, F\}$

$a_T = \{1+, 4-\} = -\frac{1}{5}\log\left(\frac{1}{5}\right) - \frac{4}{5}\log\left(\frac{4}{5}\right)$

$= 0.7219$

$a_F = \{5+, 0-\} = -\frac{5}{5}\log\left(\frac{5}{5}\right) = 0.$

$I_G(a) = 0.9709 - 0.7219 \times \frac{5}{10} - 0$

$= 0.6099$

a_2 : {HOT, COOL}

a_{HOT} : {2+, 3-} $= -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$

$= 0.9709$

a_{COOL} : {4+, 1-} $= -\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5}$

$= 0.9219.$

$I_G(a_2) = 0.9709 - \frac{5}{10}(0.9709) - \frac{5}{10} \times 0.7219$

$= 0.1245$

a_3 = {High, Normal}

a_{High} : {2+, 4-} $= \frac{-2}{6}\log\frac{2}{6} - \frac{4}{6}\log\frac{4}{6}$

$= 0.9183$

a_{Nor} : {4+, 0} : $\frac{-4}{4}\log\frac{4}{4} = 0$ ✓

$I_G(a_3) = +0.9709 - \frac{6}{10} \times 0.9183 - \frac{4}{10} \times 0.$

$= 0.4199$ ✓

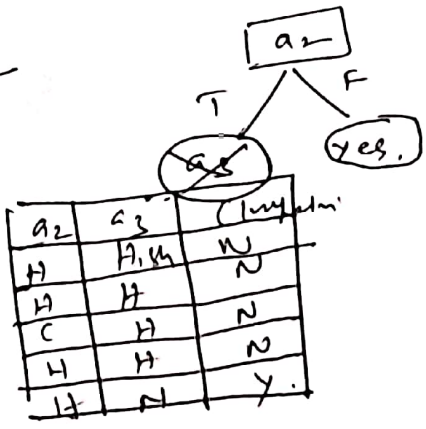
Step 2 :-

Table Entropy

= {1+, 4-}

$= -\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5}$

$= 0.7219.$



a_2	a_3	(illegible)
H	H, N	N
H	H	N
C	H	N
H	H	N
H	N	Y.

a_2 : {Hot, Cool}

E_{Hot} : {3+, 1−}

$= -3/4 \log 3/4 - 1/4 \log 1/4 = 0.8112$

E_{Cool} : {1−} $= 0$.

I_G : $0.7219 - \frac{4}{5} \times 0.8112 - 1/5 \times 0$

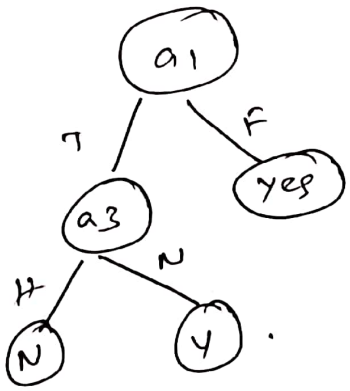
$= 0.3219$

a_3 : {High Normal}

a_{High} : {4−} $= 0$.

a_{Normal} : {1+} $= 0$.

I_G : $0.7219 - 0 =$ 0.7219



Termination Conditions of Genetic Algorithm.

Maximum Generation :-

The GA stops when the specified number of generations have been evolved.

Elapsed Time :-

The generation process will end when a specific time has been elapsed.

No Change in fitness :-

The genetic process will end if there is no change to the populations fitness for a specified number of generations.