Unit-3

Classification :-

Classification and Prediction :-

Classification :-

- The classification is an important technique of data minging.
- It is one of the most commonly used technique in data mining.
- classification technique is used for analysing the data.
- Finding a good model that is used to predict the class of objects whose class label is unknown.
- Categorization of new data with the help of current/past data.

Example:

Grouping of the patients based on their Medical Records.

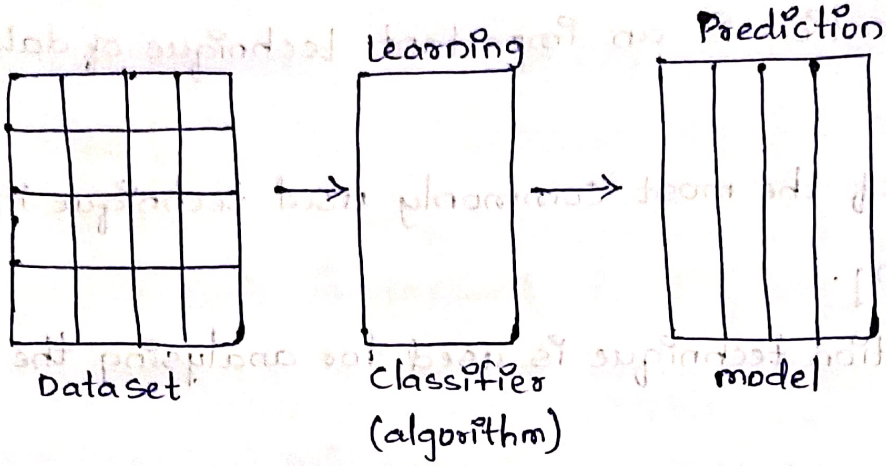
Prediction :-

- Prediction is a technique of predicting the data which have been ~~predicted~~ classified.
- Predicting a missing/unknown value base on past/current data.
- It's output is a continuous value.

Example:

Predicting the correct treatment for a person based on their medical record.

→ The below given is diagramatic representation of classification and Prediction.



Basic Concepts :-

Decision Tree Induction :-

- Decision Tree is a tree like structure.
- It supports in taking decisions.
- It defines the rules visually in the form of tree.
- In decision tree, the data is represented in the form of tree for prediction and the decision.
- The decision tree will make use of different compon-ents. The different components of decision tree are:
  - Root node - main Question
  - Leaf node - Answer
  - Internal node - Intermediate process.

Attributes selection Measure:

1. Information Gain:

How much information provided for the specific Question to the answer.

2. Entropy:

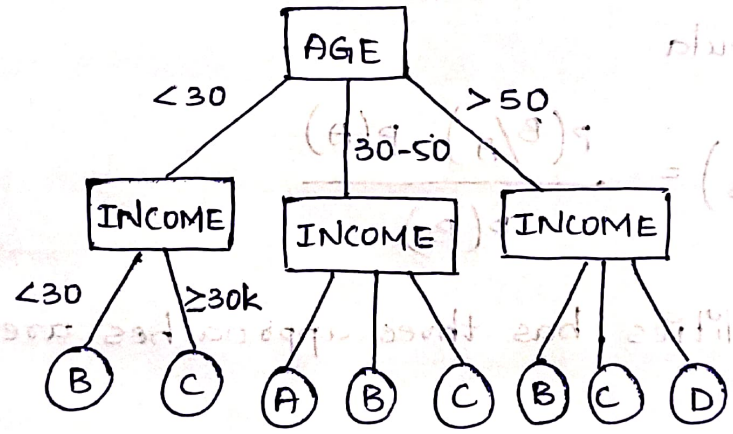
Entropy measures the amount of uncertainty in the information.

(As Information Gain increases, entropy also decreases).

Example:

credit score Rating.

A → Average, B → Bad, C → Good, D → Excellent



Rules that can be defined as

If age < 30, income < 30k, the credit score = Bad

If age ≤30, income ≥ 30k, the credit score = Good.

Bayesian classification :-

- The bayesian classifiers are based on bayesian networks.
- The bayesian networks are based on bayesian methods.
- The bayesian classifiers are meant to classify the data.
- The bayesian classifies are proposed by Thomos bayes.
- The other names of Bayesian Networks are : Bayesian belief networks, Probabilistic networks, etc.
- The Bayesian theorem indicates the ~~probability of an event~~ conditional probability and the conditional probability of an event is given by using formula.

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

- The probabilities has three approaches are as follows

1. Degree belief
2. Relative Frequency.
3. Associomatic probability.

1. Degree belief :

In this approach of probability, the uncertainity are measured based on the personal opinion.

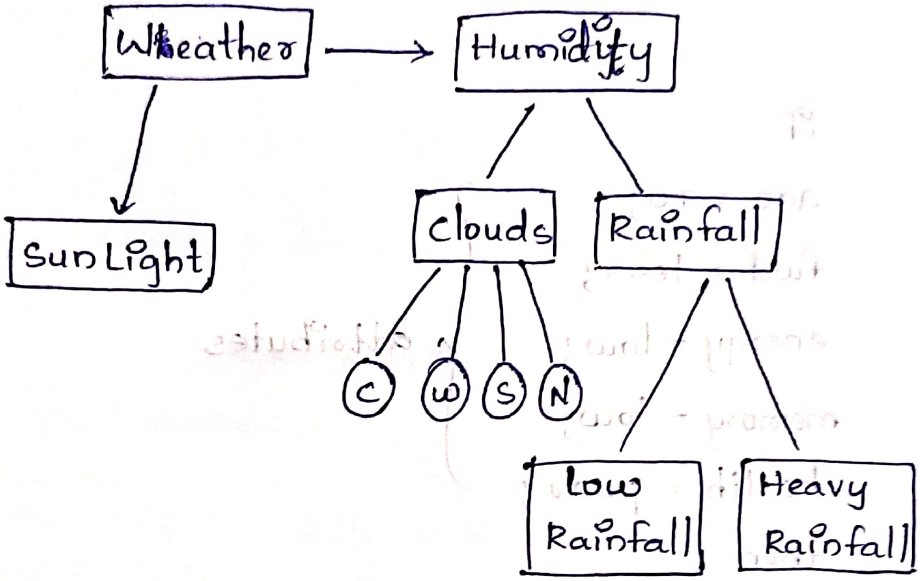
2. Relative Frequency :

In this approach of probability, the uncertainity are measured based on the trials of the data.

3. Assiomatic Poobability:

In this approach of probability, the Uncertainity are measuring by taking definitions and the properties.

- The Bayesian belief networks are easy to understand and also these classifies accurately.
- The below given is a small bayesian network.



- These are various Bayesian algorithms and some of these are Baye's net, Naive Baye's normal, Naive Baye's simple, etc.

Rule-Based Classification :-

- Rule-based classification can be done by using if then clauses.
- It uses set of IF THEN rules for classification
- There are three Important keywords

  IF, AND, THEN.

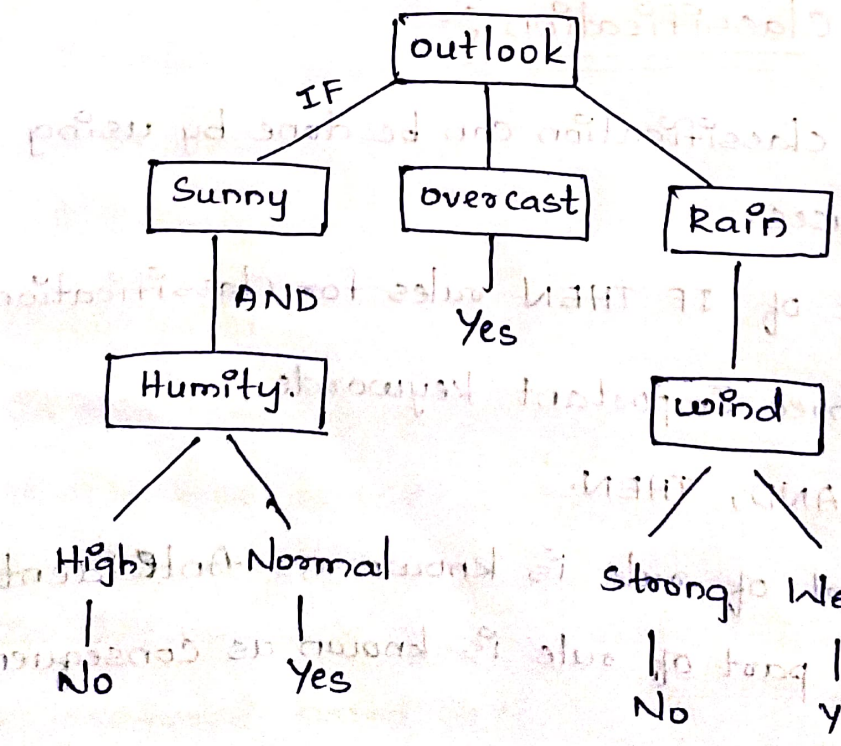
- The IF part of rule is known as Antecident and THEN part of rule is known as consequent.

- In Rule-based classification the rules are formed using by Antecedent and consequent.
- The Rule-based classification is easy to understand simple to implement.
- The Rule-based classification improves the accuracy of classification technique.
- Below given a simple rule for Rule-based classification.

```
if
age = 70;
fwd = less;
energy = low;
memory = low;
health = poor;
then
old man; } class.
```

attributes

- With the help of decision tree
  - Extract Rules from Decision tree.



Lazy Learners :-

- Lazy Learners are learning from Neighbours.
- It simply stores training data and waits to get test tuple.
- Lazy learners are the classification Algorithms.
- The earlier models for classification [Learning] like Decision trees, Neural Network, Bayesian Networks, Support Vector Machine, etc.

  Ex: KNN Algorithm.
- It works only when it gives a new example.
  - Less training time
  - More prediction time.