→ It's used for measuring temperature, pressure, location, traffic, etc---

→ Using RFID sensors we can keep track of parcels, goods & delivery.

→ Sensors are used in IOT Applications.

Q) Explain briefly about preprocessing?

A) Data is collected from various sources which contains heterogenous data.

→ Due to this the data can be inconsistent & noisy.

→ If the data is inconsistent then data mining process gives inacurate results.

→ Therefore to make data consistent, data preprocesing is required.

→ Data preprocesing improves data quality & reduces the difficulty of data mining process

Data PreProcessing Techniques :-

(i) Data cleaning
(ii) Data Integration
(iii) Data Transformation
(iv) Data Reduction

Data cleaning :- It's a process of removing unnecessary and inconsistent data from data base.
→ In data cleaning we fill the missing values to improve the quality of data

Methods for filling missing values :-
(i) Ignoring the tuples with missing values :- In this method we ignores the tuples whose values are missing.

→ This method is good when more number of values are missing in the same tuple and vice-versa

(ii) Manually filling the missing data values :- In this method the user if self find the missing values and fill them manually. It's practically impossible because of large database size

(iii) Using global constant values :- In this method we fill the missing values using global constants like "Unknown", "∞", "Null", "—". It's a simple approach but it gives in-accurate result because data mining process treat global constant specially.

(iv) Using attribute average values :- In this method the missing value is filled with the average value of that attribute.

(v) Using most likely desirable values :- This is a popular approach in which current information is used to predict the missing value.

(ii) Data Integration:- It's a process of combining data from various heterogenous data sources such as database, files, images, etc - - - -.

→ To form a single consistent data repository.

→ There are 3 problems associated with data integration

(i) Object matching & schema integration:- Object matching is a process of matching the objects based on their meaning rather than names. In schema integration the errors are avoided using metadata.

(ii) Redundancy & Inconsistency:- Redundancy is nothing but repetation of data. Redundancy makes duplicate tuples in database which makes results inaccurate.

→ Therefore it's necessary to detect and delete duplicate data.

(iii) Identifying & resolving the conflix b/w the data values:- This issue takes place when we collect objects which are similar but with different attribute values.

Ex:- In one of the databases the cost of a bicycle is stored in rupees, but in other database it's stored in dollar

(iii) Data Transformation :- This is a process of converting the integrated data into correct format for performing data mining, following are the techniques :-

(i) Smoothing of noisy data
(ii) Generalization of data
(iii) Normalization of data

Smoothing of noisy data :- These techniques are used to remove noisy data from the entire database to achieve accurate results.

Ex :- Regression, clustering

generalization of data :- In this technique all the values present at lower conceptual level are substituted by values present at higher conceptual level.

Normalization of data :- It's a process of decomposing the value to match with smaller size value.

Ex :- Min Max normalization, z-score normalization, decimal scale normalization