**DATA PREPROCESSING:**

- what is data preprocessing

- why data pre-process

- how to data pre-process

- when should one data pre-process

- Techniques of data pre-process

1. **What is data preprocessing?**

It is the aspect of transforming real world data to informative data.

World data is:-

* Messy
* Inconsistent
* Missing Values
* Containing errors

1. **Why pre-process data?**

* Data quality and accuracy to improve reliability
* Improve productivity

1. **When should one data pre-process?**

Right after acquiring the data

1. **How to data pre-process?**

Data Cleaning

Data transformation

Data Integration

Handling of Imbalanced data set

Normalization and Standardization

Feature Engineering

1. **Techniques of data pre-process**

Imputation

Deletion

Outliers

Duplicates

Scaling(Normalization, standardization & Robust scaling)

Encoding

Data integration

OverSampling/Undersampling/Synthetic Data Generation

Dimensionality Reduction:- PCA & Feature Selection

Tokenization

Stemming/Lemmatization

Feature Engineering(binning, creating new features)