

Preprocessing

What is Preprocessing ?

Preprocessing is the process of **cleaning and transforming raw data** into a form that can be effectively used by machine learning models.
which helps improve the **accuracy**, **speed**, and **efficiency** of models.

Why is Preprocessing Important?

- Removes noise, inconsistencies
- Deals with missing or null values
- Converts text/categorical data into numerical values.
- Increases model performance and reduces training errors.

Handling Missing or Null Values

Missing data (null or blank) can occur due to various reasons like data entry errors, sensor failure, or scraping issues.

By using the Pandas Library we find the these values

```
# Check total null values per column
print(dataframe.isnull().sum())

# Check percentage of missing values
print((dataframe.isnull().sum() / len(data_frame)) * 100)

# View rows with any missing value
print(dataframe[dataframe.isnull().any(axis=1)])

#Drop the null values
dropna(inplace=True)
```

By using fillna() method we can fill the null values

Fill with mean (for numeric),
mode (for categorical), or median.
Preprocessing is a **vital step** in any machine learning project.