Day-5
Encoding Categorical variable

Encoding Categorical Variables

Encoding categorical variables is the process of converting categorical data (like labels or categories) into a numerical format that machine learning models can understand. There are several types of encoding techniques:

1. Label Encoding (used only for output)

- **Description**: Converts each category into a unique integer.
- **Example**: For a "Color" feature:
    - Red → 0
    - Blue → 1
    - Green → 2
- **Use Case**: Suitable for ordinal data (where the categories have a meaningful order).

2. One-Hot Encoding

- **Description**: Creates binary columns for each category, with a 1 indicating the presence of that category and 0 indicating its absence.
- **Example**: For a "Color" feature:
    - Red → [1, 0, 0]
    - Blue → [0, 1, 0]
    - Green → [0, 0, 1]
- **Use Case**: Suitable for nominal data (where categories have no inherent order).

3. Ordinal Encoding

- **Description**: Assigns ordered integers to categories based on their rank or order.
- **Example**: For "Education Level":
    - School → 0
    - Under Graduate → 1
    - Post Graduate → 2
- **Use Case**: Used when there is a clear order among the categories.