Page No. :
Date. : / /

Q1.) What is Label Encoding?

ANS. Label Encoding is a technique used in Machine Learning to convert categorical data (text labels) into numerical values.

- Explaination:-
  - Many ML Algorithms work only with numbers.
  - So categories are assigned integer values.

→ Example:-

| Category | Encoded Value. |
|---|---|
| Red | 0 |
| Green | 1 |
| Blue | 2 |

- Features:-
  - Simple and Fast.
  - Used for ordinal Data. (where order matters)
  - Not suitable for nominal data. (no order)

→ Label encoding converts text labels → numbers for ML Models.

Q2.) What are the Lemmatization Methods? Explain any one of them?

ANS. Lemmatization is the process of reducing a word to its base / lemma form i.e. dictionary form using vocabulary or grammar rules.

Eg:-

① running → run ③ studies → study

② better → good

- It is slower but more accurate.
- Always gives a valid dictionary word.

Page No. :
Date. : / /

Lemmatization Methods:-

→ 1. Rule based Lemmatization
2. Dictionary based Lemmatization
3. Machine Learning based Lemmatization.

→ **Rule Based Lemmatization:-**

How it Works.
- Uses grammar rules and patterns.
- Considers Part of Speech (POS)

**Example:-**

"playing" → remove "ing" → "play"
"studies" → study

**Steps:-**

1. Identify Word Form
2. Apply grammar rules.
3. Convert to base Form.

**Advantages:-**

. Easy to Implement and Fast.

. **Dis-Advantages**

- Not Always Accurate
- Cannot handle irregular words well

→ Rule based lemmatization uses linguistic rules to find root words.

Q3.) What is the need of Text Cleaning? How it is done?

ANS. Need of text cleaning:-

Raw text contains noise;

. Punctuation

Page No. :
Date. : / /

- Special characters.
- Stopwords.
- Irrelevant words.

→ Importance :-

1. Improves Model Accuracy.
2. Reduces Noise.
3. Speeds Up processing.
4. Standurizes text.

⇒ How text cleaning is Done :-

1. Lowercasing
   - "Hello" → "hello"

2. Remove Punctuation
   - "hello !!!" → "hello"

3. Removing stopwords.
   - Words like : "is" ; "the", "and"

4. Tokenization
   - Split sentence into words.

5. Stemming / Lemmatization
   - Converts words to root or base form.

6. Removing special characters and Numbers.
   - "abc123@" → "abc"

7. Removing extra spaces.
   - removing white spaces.

Example :- " This is a GREAT product !!! "

↓ Text cleaning.

" great product "

→ Text cleaning prepares raw text into usable format for NLP Models.