**Training Report Day-15**

**26 June 2024**

**NLP(Natural language Processing):**

**Text processing**

* Tokenization: splitting text into individual words.
* Lowercasing: Converting all text to lower case.
* Removing punctuation: Eliminating punctuation marks.
* Removing stop words: Removing common words that may not contribute significantly to the meaning. (e.g. “and”,”the”)
* Stemming and Lemmatization:Reducing words to their root form (e.g., "running" to "run").
* Removing Numbers: Depending on the context, numbers may be removed.
* Handling Special Characters: Removing or transforming special characters (e.g., emojis, HTML tags).