

## **NLP Documentation**

Libraries Used:

- **Pillow:**

- Pillow supports the following:
  - Image resizing and cropping
  - Image filtering and enhancement
  - Image format conversion
  - Text and annotation overlay
  - Image manipulation and composition
  - Image analysis and processing
  - Image drawing and visualization

- **cv2:**

- cv2 supports the following:
  - Image and Video Processing
  - Object Detection and Tracking
  - Feature extraction and matching
  - Image Segmentation
  - Camera Calibration
  - Optical Flow and Motion Analysis
  - Deep Learning integration
  - Augmented Reality

- **EasyOCR:**

- EasyOCR supports the following:
  - Text Extraction from Images
  - Automatic Data Entry
  - Text Recognition from videos

- Language Translation
- Real-Time Text Recognition
- Document Analysis

- **NLTK:**

- Natural Language Toolkit, supports the following:
  - Text Preprocessing
  - Tokenization
  - Tagging
  - NER: Named Entity Recognition
  - Sentiment Analysis
  - Language Modelling and Generation
  - Text Classification
  - Corpus and linguistic resources

- Steps Followed in this NLP Process:

- Loaded image using **cv2**
- Extracted textual data using **EasyOCR**
- Performed Data Cleaning:
  - Removing Extra Spaces
  - Removing Symbols
  - Replacing new line tags with single spaces
  - Transformed data to lowercase
- Removed stopwords using **nltk**
- Tokenized data using **nltk**
- Tagging data using **nltk**
- Entity Extraction using **nltk**