	Bansilal Ramnath Agarwal Charitable Trust's Vishwakarma Institute of Information Technology **Department of Artificial Intelligence and Data Science**

Name: Siddhesh Dilip Khairnar		
Class: TY	Division: B	Roll No: 372028
Semester: 5	Academic Year:2023-2024	
Subject Name & Code: Natural Language Processing (ADUA32203)		
Title of Assignment: Comparative study of available libraries for Natural Language processing with respect to provided functionalities, platform dependence, supported NLP approaches, supported L P Tasks, advantages and Disadvantages etc.		
Date of Performance: 09-08-2023	Date of Submission: 10-08-2023	

Aim: To do a Comparative study of available libraries for Natural Language processing with respect to provided functionalities, platform dependence, supported NLP approaches, supported L P Tasks, advantages and Disadvantages etc.

Problem Statement: To do a Comparative study of available libraries for Natural Language processing with respect to provided functionalities, platform dependence, supported NLP approaches, supported L P Tasks, advantages and Disadvantages etc.

NLTK (Natural Language Toolkit):

(a) Functionalities: NLTK offers a wide range of NLP tools and resources, including tokenization, stemming, part-of-speech tagging, named entity recognition, sentiment analysis, and more.
(b) Platform Dependence: Cross-platform (Linux, macOS, Windows).
(c) Supported NLP Approaches: Traditional rule-based NLP approaches.
(d) Supported NLP Tasks: Comprehensive support for various NLP tasks.
(e) Advantages: Well-established, extensive documentation, educational resources.
(f) Disadvantages: Some components might be slower than alternatives for large-scale processing.

spaCy:

(a) Functionalities: Efficient tokenization, part-of-speech tagging, named entity recognition, dependency parsing, and more. Platform Dependence: Cross-platform.
(b) Supported NLP Approaches: Modern, statistical, and rule-based approaches.
(c) Supported NLP Tasks: Focus on core NLP tasks with a focus on production efficiency.
(d) Advantages: High performance, pre-trained models, easy-to-use API.
(e) Disadvantages: Limited resources for languages other than English.

TextBlob:

(a) Functionalities: Simple API for common NLP tasks, including sentiment analysis, noun phrase extraction, classification, translation, and more.
(b) Platform Dependence: Cross-platform.
(c) Supported NLP Approaches: Primarily rule based.
(d) Supported NLP Tasks: Basic NLP tasks with an emphasis on simplicity.
(e) Advantages: Easy to use, beginner friendly.
(f) Disadvantages: May lack advanced features for complex tasks.

Hugging Face Transformers:

(a) Functionalities: Provides pre-trained models for various NLP tasks, such as text classification, language modeling, translation, and more.
(b) Platform Dependence: Cross-platform.
(c) Supported NLP Approaches: Deep learning, transformer-based models.
(d) Supported NLP Tasks: Diverse tasks with a focus on state-of-the- art models.
(e) Advantages: Access to cutting-edge models, easy model deployment.
(f) Disadvantages: May require significant computational resources for large models.

scikit-learn:

(a) Functionalities: General-purpose machine learning library with some NLP modules, including text feature extraction and basic text classification.
(b) Platform Dependence: Cross-platform.
(c) Supported NLP Approaches: ML, statistical methods.
(d) Supported NLP Tasks: Basic text classification, clustering, feature extraction.
(e) Advantages: Broad ML functionality, well-documented.
(f) Disadvantages: Limited compared to specialized NLP libraries.

TensorFlow and PyTorch:

(a) Functionalities: Deep learning frameworks with extensive support for building custom NLP models.
(b) Platform Dependence: Cross-platform.
(c) Supported NLP Approaches: Deep learning, neural networks.
(d) Supported NLP Tasks: Customizable for various NLP tasks.
(e) Advantages: Flexibility, community support, scalability.
(f) Disadvantages: Steeper learning curve for beginners.

Stanford NLP:

(a) Functionalities: Suite of NLP tools, including part-of-speech tagging, named entity recognition, sentiment analysis, and more.
(b) Platform Dependence: Cross-platform.
(c) Supported NLP Approaches: Rule-based, machine learning.
(d) Supported NLP Tasks: Various NLP tasks.
(e) Advantages: Robust tools, well-established.

(f) Disadvantages: Some components might be resource intensive.

Apache OpenNLP:

(a) Functionalities: Toolkit for various NLP tasks, including tokenization, part-of-speech tagging, named entity recognition, and more.

(b) Platform Dependence: Cross-platform.

(c) Supported NLP Approaches: Machine learning based.

(d) Supported NLP Tasks: Common NLP tasks.

(e) Advantages: Well-integrated, actively developed.

(f) Disadvantages: May require more effort for certain tasks compared to newer libraries.

Library	Functionalities	Platform Dependence	Supported NLP	Supported NLP Tasks	Advantages
NLTK	Comprehensive NLP tools	Cross-platform	Traditional rule-based NLP	Various NLP tasks	Well- established, extensive documentation
spaCy	Efficient NLP tasks	Cross-platform	Modern, statistical, rule-based	Core NLP tasks with production efficiency	High performance, pre-trained
Text Blob	Simple API for common NLP tasks	Cross-platform	Primarily rule-based	Basic NLP tasks with simplicity	Easy to use, beginner-friendly
Hugging Face Transformer	Pre-trained models for various NLP tasks	Cross-platform	Deep learning, transformers	Diverse tasks with state-of-the-art models	Access to cutting- edge models, easy
scikit-learn	General- purpose ML library with some NLP	Cross-platform	Machine learning, statistical	Basic text classification, clustering	Broad ML functionality, well-documented
TensorFlow and PyTorch	Deep learning frameworks for custom	Cross-platform	Deep learning, neural networks	Customizable for various NLP tasks	Flexibility, community support, scalability
Stanford NLP	Suite of NLP tools	Cross-platform	Rule-based, machine learning	Various NLP tasks	Robust tools, well- established
Apache OpenNLP	Toolkit for various NLP tasks	Cross-platform	Machine learning-based	Common NLP tasks	Well- integrated, actively developed