Assignment: Comprehensive NLP Project Using Python

**Project Description**:

Create a unified NLP application that serves as an "Intelligent Text Analyzer." This tool will process a collection of text documents (such as news articles, blogs, or reviews) and perform multiple analyses: categorizing content, extracting key entities, assessing sentiment, and classifying documents into topics.

## Tasks:

**1. Data Collection and Preprocessing:**

* Collect a dataset suitable for all four tasks. You may use datasets like news articles, movie or product reviews.
* Preprocess the data to make it suitable for NLP tasks (tokenization, removing stopwords, etc.).

**2. Text Classification:**

* Implement a text classification model using scikit-learn to categorize text based on predefined categories (e.g., sports, politics, technology in news articles).
* Analyze the performance of the model using accuracy metrics and confusion matrix.

**3. Named Entity Recognition (NER):**

* Use spaCy to identify named entities (people, organizations, locations, etc.) in the text.
* Examine how these entities correlate with the text categories identified in the text classification task.

**4. Sentiment Analysis:**

* Implement sentiment analysis on the same set of texts using nltk.
* Determine the overall sentiment (positive, negative, neutral) of texts in different categories.

**5. Document Classification:**

* Using the same dataset, classify the documents into more specific topics or themes using an unsupervised learning approach (like LDA).
* Explore how these topics align with the categories and sentiments identified in previous tasks.

## Deliverables:

* A Python script(s) that integrates all four NLP tasks into a cohesive workflow.
* A detailed report (4-6 pages) that includes:
* Methodology: Describe your approach for each task, including data preprocessing, model selection, and implementation details.
* Results: Present the results of each task, with visualizations and a discussion on the effectiveness of the models used.
* Analysis: Offer insights on how the results of different tasks interrelate (e.g., the relation between sentiment and text categories, or how named entities are distributed across different topics).
* Challenges and Learnings: Discuss any challenges faced during the project and what you learned from addressing them.