TechDome Assignment

News Analysis Project

Assignment Documentation

Tanay Bhutada

How to Run the Program:

1.Clone the repository to your local machine:
git clone <[repository_url](https://github.com/mishop-15/News-Analysis-Project-NLP-)>

2.Navigate to the project directory:
cd news-analysis-project

3.Install the required dependencies using pip:
pip install -r requirements.txt

4.Open the main Python file (main.py or news_analysis.py, etc.) in your preferred Python IDE or text editor.

5.Modify the file to specify the input data or any other configuration options as needed.

6.Run the Python file to execute the project:
python main.py

Requirements:

The project requires the following Python packages: numpy pandas nltk scikit-learn gensim pyLDAvis lexnlp

You can install these dependencies using the requirements.txt file provided in the repository.

To install the dependencies, run the following command in your terminal or command prompt: !pip install -r requirements.txt.

Libraries Used:

spaCy:
spaCy is a library for advanced natural language processing in Python. It provides pre-trained models and tools for various NLP tasks such as tokenization, part-of-speech tagging, named entity recognition, and dependency parsing.

pyATE (Automated Term Extraction):
pyATE is a Python library for automated term extraction, which can identify key terms or phrases in text documents based on statistical and linguistic patterns.

NLTK (Natural Language Toolkit):
NLTK is a comprehensive library for natural language processing tasks, including tokenization, stemming, lemmatization, part-of-speech tagging, and sentiment analysis. It also provides access to various corpora and lexical resources.

Gensim:
Gensim is a Python library for topic modeling, document indexing, and similarity retrieval with large text corpora. It includes implementations of algorithms such as Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) for analyzing text data.

pandas:
pandas is a powerful data manipulation library in Python, providing data structures and functions for working with structured data. It is commonly used for reading, cleaning, analyzing, and manipulating tabular data.

sumy:
sumy is a library for automatic text summarization in Python. It provides various algorithms for summarizing text documents, including LSA (Latent Semantic Analysis), Luhn, Edmundson, and more.

rake-nltk:
rake-nltk is a Python library for keyword extraction from text documents using the Rapid Automatic Keyword Extraction (RAKE) algorithm. It can identify key phrases or terms in text based on frequency and co-occurrence patterns.

WordNet:
WordNet is a lexical database of English words and their semantic relationships. It is often used for tasks such as synonymy detection, hyponymy/hypernymy extraction, and word sense disambiguation.

VADER (Valence Aware Dictionary and sentiment Reasoner):
VADER is a lexicon and rule-based sentiment analysis tool specifically designed for social media texts. It can analyze the sentiment of text data and provide sentiment scores for individual sentences or documents.

SciPy:
SciPy is a library for scientific and technical computing in Python. It includes modules for optimization, integration, interpolation, linear algebra, and more, which can be useful for various data analysis tasks.

These libraries provide a wide range of functionalities for processing, analyzing, and extracting insights from text data, making them essential tools for the News Analysis Project.

: