jupyter Untitled6 Last Checkpoint: 36 seconds ago

File Edit View Run Kernel Settings Help

Code

•[2]:

```
#LMS DAY-7(197)
!pip install gensim nltk spacy
import re
import gensim
from nltk.stem.porter import PorterStemmer
from nltk.corpus import stopwords
import spacy
import nltk
nltk.download('stopwords')
nlp = spacy.load("en_core_web_sm")
porter_stemmer = PorterStemmer()
stop_words = set(stopwords.words('english'))
def preprocess_text(text):
    text = re.sub(r'[^\w\s]', '', text.lower())
    tokens = [word for word in gensim.utils.simple_preprocess(text) if word not in stop_
    stemmed_tokens = [porter_stemmer.stem(token) for token in tokens]
    doc = nlp(' '.join(stemmed_tokens))
    lemmatized_tokens = [token.lemma_ for token in doc]
    return lemmatized_tokens
text_content = """
Write a Python script that uses Gensim to preprocess data from a sample text
file. Follow basic procedures like tokenization, stemming, and lemmatization.
Print the final output to verify the preprocessing steps.
"""
processed_text = preprocess_text(text_content)
print(processed_text)
```