```
# [Day-7 2211cs020206] Write a python script that:1.Use genism to preprocess data from a sample text file,follow basic procedure like tokenization,stemming,lemmatization
```

In [*]:

```
!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
    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
file. Follow basic procedures like tokenization, stemming, and lemmatiz
Print the final output to verify the preprocessing steps.
"""
processed_text = preprocess_text(text_content)
print(processed_text)
```

```
ERROR: Could not install packages due to an OSError: [WinError 5] Access i
s denied: 'C:\\Users\\nihar\\Anaconda3\\Lib\\site-packages\\~umpy.libs\\li
bscipy_openblas64_-caad452230ae4ddb57899b8b3a33c55c.dll'
Consider using the `--user` option or check the permissions.
```

```
Requirement already satisfied: gensim in c:\users\nihar\anaconda3\lib\site
-packages (4.1.2)
Requirement already satisfied: nltk in c:\users\nihar\anaconda3\lib\site-p
ackages (3.7)
Requirement already satisfied: spacy in c:\users\nihar\anaconda3\lib\site-
packages (3.7.2)
Requirement already satisfied: numpy>=1.17.0 in c:\users\nihar\anaconda3\l
ib\site-packages (from gensim) (2.0.2)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\nihar\anacond
a3\lib\site-packages (from gensim) (5.2.1)
Requirement already satisfied: scipy>=0.18.1 in c:\users\nihar\anaconda3\l
ib\site-packages (from gensim) (1.9.1)
Requirement already satisfied: joblib in c:\users\nihar\anaconda3\lib\site
-packages (from nltk) (1.1.0)
Requirement already satisfied: click in c:\users\nihar\anaconda3\lib\site-
packages (from nltk) (8.0.4)
Requirement already satisfied: regex>=2021.8.3 in c:\users\nihar\anaconda3
\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\nihar\anaconda3\lib\site-p
ackages (from nltk) (4.64.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\nihar\a
naconda3\lib\site-packages (from spacy) (2.0.10)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in
c:\users\nihar\anaconda3\lib\site-packages (from spacy) (2.5.2)
Requirement already satisfied: packaging>=20.0 in c:\users\nihar\anaconda3
\lib\site-packages (from spacy) (21.3)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\nihar\anaco
nda3\lib\site-packages (from spacy) (2.0.8)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\nihar\a
naconda3\lib\site-packages (from spacy) (2.28.1)
Requirement already satisfied: jinja2 in c:\users\nihar\anaconda3\lib\site
-packages (from spacy) (2.11.3)
Requirement already satisfied: thinc<8.3.0,>=8.1.8 in c:\users\nihar\anaco
nda3\lib\site-packages (from spacy) (8.2.1)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\nihar
\anaconda3\lib\site-packages (from spacy) (1.0.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\nihar\anac
onda3\lib\site-packages (from spacy) (1.1.2)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in c:\users\nihar\anac
onda3\lib\site-packages (from spacy) (0.9.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\nihar\ana
conda3\lib\site-packages (from spacy) (3.0.9)
Requirement already satisfied: setuptools in c:\users\nihar\anaconda3\lib
\site-packages (from spacy) (63.4.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\nihar\a
naconda3\lib\site-packages (from spacy) (3.3.0)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\nihar\anaco
nda3\lib\site-packages (from spacy) (2.4.8)
Requirement already satisfied: weasel<0.4.0,>=0.1.0 in c:\users\nihar\anac
onda3\lib\site-packages (from spacy) (0.3.4)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\nih
ar\anaconda3\lib\site-packages (from spacy) (1.0.5)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\nih
ar\anaconda3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\nihar
\anaconda3\lib\site-packages (from packaging>=20.0->spacy) (3.0.9)
Requirement already satisfied: annotated-types>=0.4.0 in c:\users\nihar\an
aconda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spac
y) (0.6.0)
Requirement already satisfied: typing-extensions>=4.6.1 in c:\users\nihar
\anaconda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->s
pacy) (4.8.0)
Requirement already satisfied: pydantic-core==2.14.5 in c:\users\nihar\ana
conda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4->spac
y) (2.14.5)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nihar\ana
conda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\nihar
\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\nihar\anaconda3\li
b\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\nihar\anacon
da3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.9.14)
Collecting numpy>=1.17.0
  Downloading numpy-1.24.4-cp39-cp39-win_amd64.whl (14.9 MB)
     ---------------------------------------- 14.9/14.9 MB 6.2 MB/s eta 0:
00:00
Requirement already satisfied: blis<0.8.0,>=0.7.8 in c:\users\nihar\anacon
da3\lib\site-packages (from thinc<8.3.0,>=8.1.8->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\nihar
\anaconda3\lib\site-packages (from thinc<8.3.0,>=8.1.8->spacy) (0.1.4)
Requirement already satisfied: colorama in c:\users\nihar\anaconda3\lib\si
te-packages (from tqdm->nltk) (0.4.6)
Requirement already satisfied: cloudpathlib<0.17.0,>=0.7.0 in c:\users\nih
ar\anaconda3\lib\site-packages (from weasel<0.4.0,>=0.1.0->spacy) (0.16.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\nihar\anaconda
3\lib\site-packages (from jinja2->spacy) (2.0.1)
Installing collected packages: numpy
  Attempting uninstall: numpy
    Found existing installation: numpy 2.0.2
    Uninstalling numpy-2.0.2:
      Successfully uninstalled numpy-2.0.2
```

In [ ]: