

**For a given text, perform word and sentence tokenization. Remove the stop words from the given text and create n-grams for the value of n=2.**

```
#install package nltk

from nltk.corpus import stopwords
import nltk
nltk.download('punkt_tab')
from nltk.tokenize import sent_tokenize, word_tokenize
text1 = "The data set given satisfies the requirement for model generation.
This is used in Data Science Lab"

print("\nSentence Tokenize:", sent_tokenize(text1))
print("\nWord Tokenize:", word_tokenize(text1))
print("\nStopwords:", stopwords.words('english'))

text = word_tokenize(text1)
text= [word for word in text if word not in stopwords.words('english')]
print("\nText after removing stopwords:", text)

temp=zip(*[text[i:] for i in range(0,2)])
ans=[' '.join(ngram) for ngram in temp]
print("\n2-grams:", ans)
```