EXPERIMENT 7 :: EXPLAIN ABOUT N-GRAM MODEL.

INTRODUCTION ::

N-grams are continuous sequences of words or symbols, or tokens in a document. In technical terms, they can be defined as the neighboring sequences of items in a document. They come into play when we deal with text data in NLP (Natural Language Processing) tasks. They have a wide range of applications, like language models, semantic features, spelling correction, machine translation, text mining, etc.

PROGRAM ::

```
import re

from nltk.util import ngrams

txt="Deep Learning is subset of ML and ML is subset of AI" " and these are emerging
technologies"

tokens=[token for token in txt.split(" ")]

op=list(ngrams(tokens,2))

print(op)

print("\n")

op1=list(ngrams(tokens,4))

print(op1)
```

OUTPUT ::

```
----------------N-GRAMS WITH 2 SETS-----------------------
[('Deep', 'Learning'), ('Learning', 'is'), ('is', 'subset'), ('subset',
'of'), ('of', 'ML'), ('ML', 'and'), ('and', 'ML'), ('ML', 'is'), ('is',
'subset'), ('subset', 'of'), ('of', 'AI'), ('AI', 'and'), ('and',
'these'), ('these', 'are'), ('are', 'emerging'), ('emerging',
'technologies')]


----------------N-GRAMS WITH 4 SETS-----------------------
[('Deep', 'Learning', 'is', 'subset'), ('Learning', 'is', 'subset', 'of'),
('is', 'subset', 'of', 'ML'), ('subset', 'of', 'ML', 'and'), ('of', 'ML',
```

'and', 'ML'), ('ML', 'and', 'ML', 'is'), ('and', 'ML', 'is', 'subset'), ('ML', 'is', 'subset', 'of'), ('is', 'subset', 'of', 'AI'), ('subset', 'of', 'AI', 'and'), ('of', 'AI', 'and', 'these'), ('AI', 'and', 'these', 'are'), ('and', 'these', 'are', 'emerging'), ('these', 'are', 'emerging', 'technologies')]