**Program 1.2: Advanced Frequency Distribution with Stopword Removal**

**Aim:**

To plot a frequency distribution after removing stopwords.

**Procedure**:

1. Import required libraries: nltk, matplotlib.pyplot.

2. Download and load the list of English stopwords.

3. Tokenize the input text using TreebankWordTokenizer

4. Remove stopwords from the tokenized words.

5. Calculate word frequencies using FreqDist.

6. Plot the frequency distribution using matplotlib.

**Code:**

import nltk

from nltk.tokenize import TreebankWordTokenizer

from nltk.corpus import stopwords

from nltk.probability import FreqDist

import matplotlib.pyplot as plt

*# Download stopwords set (if not already downloaded)*

nltk.download('stopwords')

*# Initialize tokenizer and stopwords list*

tokenizer = TreebankWordTokenizer()

stop\_words = set(stopwords.words('english'))

*# Sample text*

text = "Natural language processing with Python is interesting and useful. Python makes NLP

tasks easier and more efficient."

*# Tokenize the text*

tokens = tokenizer.tokenize(text)

*# Remove stopwords*

filtered\_tokens = [word for word in tokens if word.lower() not in stop\_words and word.isalpha()]

*# Print filtered tokens*

print("Filtered Tokens (without stopwords):", filtered\_tokens)

**Output**

****

*# Frequency Distribution*

fdist = FreqDist(filtered\_tokens)

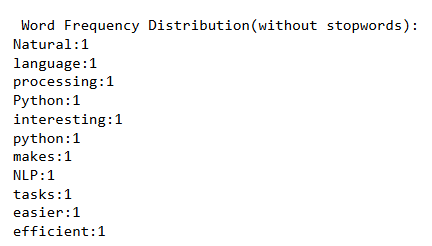
*# Print frequency distribution*

print("\nWord Frequency Distribution (without stopwords):")

for word, freq in fdist.items():

print(f"{word}: {freq}")

**Output:**

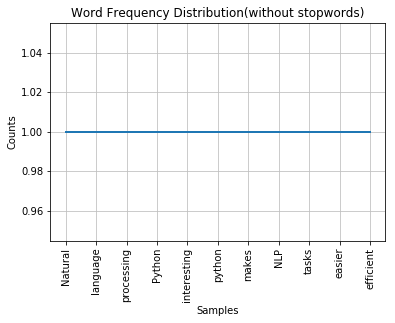
****

*# Plot frequency distribution*

fdist.plot(title='Word Frequency Distribution (Without Stopwords)')

plt.show()

**Output:**

****

*#Histogram*

plt.figure(figsize=(15,8))

plt.hist(filtered\_tokens,bins=5)

plt.show()

**Output**

****

**Result:**

Frequency distribution plotted after removing stopwords.