

Program:

```
from nltk.corpus import stopwords

def tokenize(s):
    #split by white text
    space_split = list(s.split())

    #dealing with punctuations
    punctuation = [',', '.', '!', '?']

    tokens=[]
    for i in space_split:
        f=0
        for j in punctuation:
            if j in i:
                tokens.append(i.replace(j, ""))
                tokens.append(j)
            f=1
        if f==0:
            tokens.append(i)

    #dealing with 'm and 's
    tokens2 = []
    for i in tokens:
        if '\s' in i:
            tokens2.append(i.replace('\s', ""))
            tokens2.append('\s')
        elif '\m' in i:
            tokens2.append(i.replace('\m', ""))
            tokens2.append('am')
        else:
            tokens2.append(i)
    for i in tokens2:
        if '-' in i:
            a,b = i.split('-',1)
            tokens2.append(a)
            tokens2.append('-')

    tokens2.append(b)

    return tokens2

def remove_stop_words(a):
    stop_words =
    set(stopwords.words('english'))
    b = [w for w in a if w not in stop_words]
    return b

def stemming(s):
    rem = ['ing', 'ed', 'ly', 'ive', 'able']
    b=[]
    for i in s:
        f=0
        for j in rem:
            if j in i:
                b.append(i.replace(j, ""))
                f=1
        if f==0:
            b.append(i)
    return b

with open('text.txt', 'r+') as f:
    data=f.read()
    print(data)
    tokens = tokenize(data)
    print("\nTokens: ",tokens)

    ##stop-words removal
    filtered_words =
    remove_stop_words(tokens)
    print("\nFiltered Words: ", filtered_words)

    stem = stemming(filtered_words)
    print("\nAfter Stemming: ", stem)
```

Output:

This is a sample sentence, showing off the stop words filtration.

I'm rey's friend. I worked hard.

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation.

Tokens: ['This', 'is', 'a', 'sample', 'sentence', ',', 'showing', 'off', 'the', 'stop', 'words', 'filtration', '.', 'I', 'am', 'rey', "'s", 'friend', '.', 'I', 'worked', 'hard', '.', 'Python', 'is', 'a', 'high-level', ',', 'interpreted', ',', 'interactive', 'and', 'object-oriented', 'scripting', 'language', '.', 'Python', 'is', 'designed', 'to', 'be', 'highly', 'readable', '.', 'It', 'uses', 'English', 'keywords', 'frequently', 'where', 'as', 'other', 'languages', 'use', 'punctuation', '.', 'high', '-', 'level', 'object', '-', 'oriented']

Filtered Words: ['This', 'sample', 'sentence', ',', 'showing', 'stop', 'words', 'filtration', '.', 'I', 'rey', "'s", 'friend', '.', 'I', 'worked', 'hard', '.', 'Python', 'high-level', ',', 'interpreted', ',', 'interactive', 'object-oriented', 'scripting', 'language', '.', 'Python', 'designed', 'highly', 'readable', '.', 'It', 'uses', 'English', 'keywords', 'frequently', 'languages', 'use', 'punctuation', '.', 'high', '-', 'level', 'object', '-', 'oriented']

After Stemming: ['This', 'sample', 'sentence', ',', 'show', 'stop', 'words', 'filtration', '.', 'I', 'rey', "'s", 'friend', '.', 'I', 'work', 'hard', '.', 'Python', 'high-level', ',', 'interpret', ',', 'interact', 'object-orient', 'script', 'language', '.', 'Python', 'design', 'high', 'read', '.', 'It', 'uses', 'English', 'keywords', 'frequent', 'languages', 'use', 'punctuation', '.', 'high', '-', 'level', 'object', '-', 'orient']