1.Merge.py

```
some_list = ["first_name", "last_name", "age", "occupation"]
some_tuple = ("John", "Holloway", 35, "carpenter")
# using zip() in dict()
# result=dict(zip(some_list,some_tuple))
# print(result)


# using for loop
result = {}
for i in range (0, len(some_list)):
    result[some_list[i]] = some_tuple[i]

print(result)
```

Output:

```
{'first_name': 'John', 'last_name': 'Holloway, 'age': 35,
occupation': 'carpenter'}
```

2.Stop_word_removal.py

```
from nltk.tokenize import sent_tokenize, word_tokenize
data = "All work and no play makes jack a dull boy, all work and no play"

#stop word removal
from nltk.corpus import stopwords # We imported auxiliary corpus
# provided with NLTK
stopWords = set(stopwords.words('english')) # a set of English stopwords
words = word_tokenize(data.lower())
wordsFiltered = []
filtered_stop = []
for w in words:
    if w not in stopWords:
        wordsFiltered.append(w)
    else:
        filtered_stop.append(w)

print("number of stop words",len(filtered_stop)) # Print the number of
stopwords
print(filtered_stop) # Print the stopwords
print(wordsFiltered) #Print the filtered words
```

Output:

```
[1]: runfile('C:/Users/Admin/Desktop/Swarupa/Exp
1/stop_word_removal.py', wdir='C:/Users/Admin/Desktop/Swarupa/Exp 1')
C:\ProgramData\Anaconda3\lib\site-packages\scipy\__init__.py:138:
```

```
UserWarning: A NumPy version 1.16.5 and <1.23.0 is required for this
version of Scipy (detected version 1.23.1) warnings.warn(f"A NumPy
version (np_minversion} and <{np_maxversion} is required for this
version of " number of stop words 7
 ['all', 'and', 'no', 'a', 'all', 'and', 'no']
['work', 'play', makes, jack', 'dull', 'boy', ',', 'work', 'play']
```

3.Explore_corpus.py

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

words = ["shamed","gaming","brought","filtered"]
ps = PorterStemmer()
#for words
for word in words:
    print(ps.stem(word))
#for sents
sentence = "gaming, the gamers play games"
words = word_tokenize(sentence)
for word in words:
    print(word + ":" + ps.stem(word))
```

Output:

```
C:\ProgramData\Anaconda3\lib\site-packages\scipy\__init__.py:138:
UserWarning: A NumPy version

>=1.16.5 and <1.23.0 is required for this version of Scipy (detected
version 1.23.1)

warnings.warn(f"A NumPy version >{np_minversion) and <{np_maxversion)
is required for this

version of
shame game
brought filter
gaming: game
the: the
gamers:gamer
play:play
games: game
```

4.Movie_rev_classifier.py

```
from nltk import FreqDist, NaiveBayesClassifier
from nltk.corpus import movie_reviews
```

```
from nltk.classify import accuracy
import random
documents = [(list(movie_reviews.words(fileid)), category)


for category in movie_reviews.categories()
    for fileid in movie_reviews.fileids(category)]
random.shuffle(documents) # This line shuffles the order of the documents
all_words = FreqDist(w.lower() for w in movie_reviews.words())
word_features = list(all_words)[:2000]
def document_features(document):
    document_words = set(document)
    features = {}
    for word in word_features:
        features['contains({})'.format(word)] = (word in document_words)
    return features
featuresets = [(document_features(d), c) for (d,c) in documents]
train_set, test_set = featuresets[100:], featuresets[:100]
classifier = NaiveBayesClassifier.train(train_set)
print(accuracy(classifier, test_set))
```

Output:

```
1/movie_rev_classifier.py', wdir='C:/Users/Admin/Desktop/Swarupa/
C:\ProgramData\Anaconda3\lib\site-packages\scipy\___init__.py:138:
UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for
this version of Scipy (detected version 1.23.11 warnings.warn(f"A
NumPy version >={np_minversion) and <{np_maxversion} is required for
this version of "

0.73
```