

## AI Lab\Exp\_11\TextClassification.py

```
1 import nltk
2 from nltk.stem import WordNetLemmatizer
3 from nltk.corpus import wordnet
4
5 nltk.download('punkt', quiet=True)
6 nltk.download('wordnet', quiet=True)
7 nltk.download('averaged_perceptron_tagger', quiet=True)
8 nltk.download('omw-1.4', quiet=True)
9
10 def get_wordnet_pos(tag):
11     if tag.startswith('J'):
12         return wordnet.ADJ
13     elif tag.startswith('V'):
14         return wordnet.VERB
15     elif tag.startswith('N'):
16         return wordnet.NOUN
17     elif tag.startswith('R'):
18         return wordnet.ADV
19     else:
20         return wordnet.NOUN
21
22 def lemmatize_text(text):
23     lemmatizer = WordNetLemmatizer()
24     tokens = nltk.word_tokenize(text)
25     pos_tags = nltk.pos_tag(tokens)
26
27     lemmatized = [lemmatizer.lemmatize(word, get_wordnet_pos(tag)) for word, tag in
28     pos_tags]
29     return lemmatized
30
31 if __name__ == "__main__":
32     sample_text = "The striped bats are hanging on their feet for best"
33
34     print("Original text:", sample_text)
35     print("Lemmatized:", lemmatize_text(sample_text))
36
37 test_words = ["running", "ran", "runs", "better", "best", "geese", "mice"]
38 lemmatizer = WordNetLemmatizer()
39
40     print("\nWord lemmatization examples:")
41     for word in test_words:
42         print(f"{word} -> {lemmatizer.lemmatize(word, wordnet.VERB)}")
```