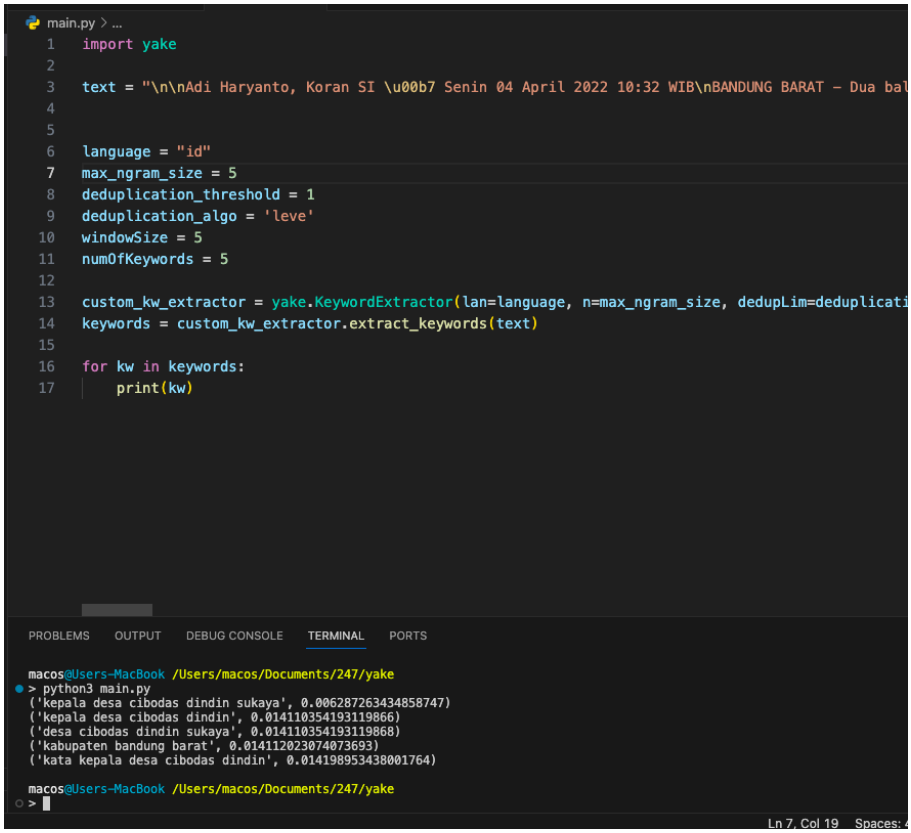
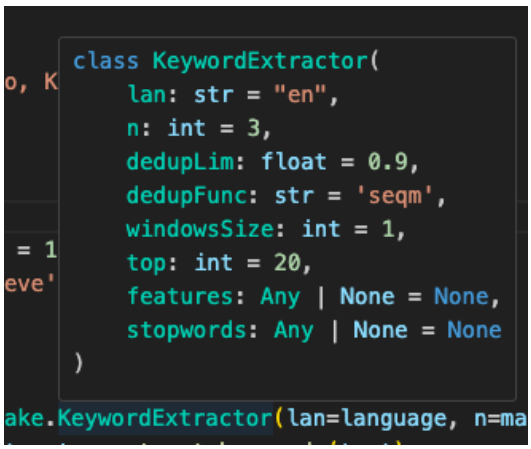
Source belajar:

- Github
- Medium
- Youtube-1
- Youtube-2

YAKE merupakan keyword extractor, yake lebih ringan dibanding rake. Parameter YAKE lebih sedikit dibanding RAKE.





Paramater Rake terdiri dari:

- lan (language),

- n (maximal panjang kata),
- deduplim (threshold untuk duplikat),
- dedubFun (algo untuk duplikat, ada ‘scm’, ‘leve’, dan ‘jargo’),
    - Sequence Matching (sqm): Algoritma ini bekerja dengan membandingkan urutan token (kata-kata) dalam kata kunci. Ini menggunakan pendekatan matching substring untuk menentukan sejauh mana suatu kata kunci mirip dengan kata kunci lainnya.
    - Levenshtein (leve): Mengukur jarak atau perubahan minimum yang diperlukan untuk mengubah satu string menjadi string lainnya. Semakin kecil jarak Levenshtein, semakin mirip kedua string tersebut.
    - Jaro-Winkler (jargo): Mengukur kemiripan antara dua string dengan memperhitungkan kesamaan karakter dan posisi karakter yang cocok. Jaro-Winkler memberikan bobot lebih tinggi pada awalan yang sama antara dua string.
- windowsSize (seberapa besar konteks yang dipertimbangkan oleh algoritma penghapusan kata-kata duplikat.),
- top (top keyword yang dihasilkan),
- stopwords(kata yang tidak bermakna)

Berbeda dengan RAKE, pada YAKE scoring nya semakin kecil maka semakin relevan