

Assignment modul 4-6

Kerjakanlah soal berikut dan kumpulkan tugas dalam 1 file notebook.

1. Sentiment Analysis

- a. Buatlah sentiment analysis menggunakan dataset IMDB_Dataset (IMDB_Dataset.csv) berdasarkan nilai polaritas.
 - i. Load dataset menggunakan read_csv, kemudian bentuk menjadi dataframe.
 - ii. Buat sentiment analysis berdasarkan kolom review.
 - iii. Perhatikan notebook pada resources modul 4 untuk cara mengerjakan.
 - iv. Dari nilai polaritas tentukan apakah suatu review positive atau negative.
- b. Boleh menggunakan textblob maupun vader (salah satu saja).
- c. Bandingkanlah hasil sentiment analysis dengan ground truth (kolom sentiment). Cukup lakukan secara manual seperti tulis kesimpulan dalam markdown/comment, atau dapat tambahkan kolom dalam dataframe yang berisi true/false.

2. Topic Modelling

- a. Lakukanlah topic modelling terhadap dataset topicmodelling.csv
 - i. Load dataset menggunakan read_csv, kemudian bentuk menjadi dataframe.
 - ii. Lakukan topic modelling berdasarkan kolom abstract.
 - iii. Perhatikan notebook pada resources modul 5 untuk cara mengerjakan.
 - iv. Cobalah beberapa num_topics, kemudian coba menggunakan 5 num_topics
 - v. Bandingkan dengan table dataset awal apakah pengelompokan topik yang didapat sama dengan pada dataset (mis. Paper a,b, e masuk ke topik 1, dll). Topik pada table bisa dilihat pada kolom Computer Science, Physics, Mathematics, Statistics, Quantitative Biology, dimana angka 1 menandakan suatu paper masuk ke topik kolom tersebut.
 - vi. Perhatikan juga apakah top words yang dihasilkan sesuai dengan topik-topik yang ada (Urutan topik bisa berbeda, Computer Science belum tentu menjadi topik pertama, dst.).

3. Text Generation

- a. Buatlah sebuah paragraph baru dengan generate text berdasarkan dataset game_of_thrones.txt
 - i. Load dataset dengan:

```
filename = '../input/game_of_thrones.txt'
data = open(filename).read()
```

(Ingat: input text generation cukup dalam bentuk string)

- ii. Perhatikan notebook pada resources modul 6 untuk cara mengerjakan.
- iii. Buatlah sebuah paragraph baru dengan 10 kalimat (lakukan 10 kali looping saat generate sentence) dengan tiap kalimat terdiri dari 15 kata.
- iv. Bacalah hasil text yang dihasilkan.