

**Kelompok:**

**Aufa Aulia Fadila Yusuf (15523070)**

**Royan Abida Nur Nayoan (15523084)**

## **Cara menghitung bobot kata dengan TF-IDF**

### **1. Term Frequency (TF)**

Menghitung frekuensi kemunculan suatu term/kata pada dokumen yang bersangkutan.

Apabila nilai TF tinggi dalam dokumen, bobot /nilai kesesuaian akan semakin besar.

- a. **TF biner (binary TF)** : Melihat ada tidaknya kata dalam dokumen. Dengan memberi nilai 1 jika ada, dan 0 jika tidak ada.
- b. **TF murni (raw TF)** : Berdasarkan kemunculan kata dalam dokumen.
- c. **TF logaritmik** dihitung dengan :

$$TF = \begin{cases} 1 + \log_{10}(f_{t,d}), & f_{t,d} > 0 \\ 0, & f_{t,d} = 0 \end{cases}$$

$f_{t,d}$  (frekuensi kata pada dokumen)

- d. **TF normalisasi** dihitung dengan :

$$TF = 0.5 + 0.5 \times \left[ \frac{f_{t,d}}{\max\{f_{t',d} : t',d \in d\}} \right]$$

### **2. Inverse Document Frequency (IDF)**

Menghitung kemunculan kata di seluruh dokumen. IDF akan semakin besar apabila jumlah dokumen yang mengandung kata yang dimaksud sedikit.

$$IDF_j = \log(D/df_j)$$

D (jumlah semua dokumen)

df<sub>j</sub> (jumlah dokumen yang mengandung kata)

Perhitungan TF yang biasa digunakan adalah TF murni, sehingga rumus TF-IDF adalah :

$$w_{ij} = tf_{ij} \times idf_j$$

$$w_{ij} = tf_{ij} \times \log(D/df_j)$$

W<sub>ij</sub> (bobot kata terhadap dokumen)

Apabila D = df<sub>j</sub> maka hasilnya 0. untuk itu ditambahkan 1 pada perhitungan IDF, sehingga menjadi :

$$w_{ij} = tf_{ij} \times \log(D/df_j) + 1$$

## **Tujuan dan manfaat penggunaan TF-IDF dalam pengolahan bahasa alami**

### **1. Tujuan**

- Untuk mendapatkan informasi (Information Retrieval).
- Text mining.

### **2. Manfaat**

- Mengevaluasi kata yang penting dalam dokumen / kumpulan dokumen.
- Mendapatkan informasi penting dari dokumen yang heterogen

## **Contoh implementasi penggunaan TF-IDF di dalam pengolahan bahasa alami**

1. Mengekstraksi ringkasan sentimen positif dan negatif masyarakat terhadap topik selebriti di Twitter..
2. Analisis tren penelitian.
3. Klasifikasi SMS.

## Referensi

<https://informatikalogi.com/term-weighting-tf-idf/>

<https://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html>

<http://www.tfidf.com/>

<https://www.cnnindonesia.com/teknologi/20181011081509-185-337518/grab-bakal-blokir-a-kun-konsumen-yang-sering-cancel-pesanan>

Al-Talib, G., & Hassan, H. (2013). A Study on Analysis of SMS Classification Using TF-IDF Weighting. *International Journal of Computer Networks and Communications Security*, 1(5), 189–194. Retrieved from [http://www.ijcnscs.org/published/volume1/issue5/p3\\_1-5.pdf](http://www.ijcnscs.org/published/volume1/issue5/p3_1-5.pdf)

Oh, K., Lim, C. G., Kim, S. S., & Choi, H. J. (2013). Research trend analysis using word similarities and clusters. *International Journal of Multimedia and Ubiquitous Engineering*, 8(1), 185–196.

Wahid, D. H., & SN, A. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(2), 207. <https://doi.org/10.22146/ijccs.16625>