**Kelompok:**

**Aufa Aulia Fadila Yusuf (15523070)**

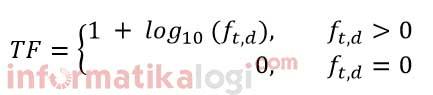
**Royan Abida Nur Nayoan (15523084)**

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

1. **Term Frequency (TF)**

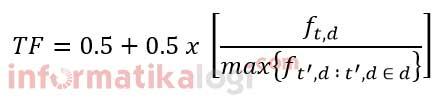
Menghitung frekuensi kemunculan suatu term/kata pada dokumen yang bersangkutan. Apabila nilai TF tinggi dalam dokumen, bobot /nilai kesesuaian akan semakin besar.

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



ft,d (frekuensi kata pada dokumen)

1. **TF normalisasi** dihitung dengan :



1. **Inverse Document Frequency (IDF)**

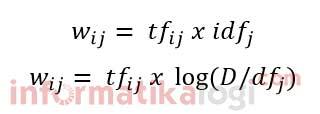
Menghitung kemunculan kata di seluruh dokumen. IDF akan semakin besar apabila jumlah dokumen yang mengandung kata yang dimaksud sedikit.

idf - inverse document frequency

D (jumlah semua dokumen)

dfj (jumlah dokumen yang mengandung kata)

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



Wij (bobot kata terhadap dokumen)

Apabila D = dfj maka hasilnya 0. untuk itu ditambahkan 1 pada perhitungan IDF, sehingga menjadi :

term weighting tf idf +1

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

1. Tujuan

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

1. 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-akun-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.ijcncs.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