**Tugas Mata Kuliah Pengolahan Bahasa Alami**

15523083 – Kartika Rizqi Nastiti

15523198 – Anisa Miladya Hakim

**Ringkasan TF-IDF (*Term Frequent – Inverse Data Frequency*)**

1. **Cara Menghitung**

TF (*Term Frequency*) digunakan untuk menghitung frekuensi kemunculan kata di setiap dokumen di dalam korpus. Cara memperoleh nilai TF adalah dengan membagi banyaknya kata terpilih dalam dokumen dengan total banyak kata di dalam dokumen tersebut. Setiap dokumen memiliki nilai TF masing-masing.

Sedangkan IDF digunakan untuk menghitung bobot kata yang jarang muncul di semua dokumen dalam korpus. Kata yang jarang muncul di dalam korpus memiliki nilai IDF yang tinggi. IDF akan melakukan proses *scaling* pada TF. Sebuah *term* yang jarang ditemui pada koleksi dokumen atau bisa dikatakan sebagai *term* khusus akan memiliki nilai IDF yang tinggi. Cara memperoleh nilai IDF adalah dengan mencari nilai *log* dari pembagian antara banyak dokumen dengan seberapa banyak sebuah kata muncul dari keseluruhan dokumen.

Nilai TF-IDF didapatkan dengan melakukan perkalian antara TF dan IDF.

1. **Tujuan dan Manfaat**
2. **Tujuan**

* Mendapatkan nilai frekuensi kemunculan suatu kata dari keseluruhan kata di dalam dokumen.
* Menghitung bobot kata yang jarang muncul di semua dokumen dalam korpus.

1. **Manfaat**

* Mengetahui kata yang paling sering muncul dari sebuah dokumen (pencarian kata kunci).
* Menentukan relevansi dokumen dengan sebuah *query* (*term*).

1. **Contoh Implementasi dalam Pengolahan Bahasa Alami**

* Analisis Sentimen
* Klasifikasi Topik
* *Automatic Text Summarization*
* Pencarian Teks

Referensi :

<https://medium.freecodecamp.org/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a94a3>

<http://repository.usu.ac.id/bitstream/handle/123456789/63920/Chapter%20II.pdf?sequence=3&isAllowed=y>

<https://prasetiautamacv.wordpress.com/2016/07/31/tf-idf-vsm-menggunakan-python/>