

Nama : Akmal Zuhdy Prasetya
NIM : H071191035
Kelas : Pengantar Data Mining

Resume Pertemuan 7

A. Thresholding Data Citra

Thresholding pada data citra adalah metode pemrosesan gambar yang membuat gambar bitonal (biner) berdasarkan pengaturan nilai ambang batas pada intensitas piksel gambar asli. Meskipun paling sering diterapkan pada gambar skala abu-abu, itu juga dapat diterapkan pada gambar berwarna.

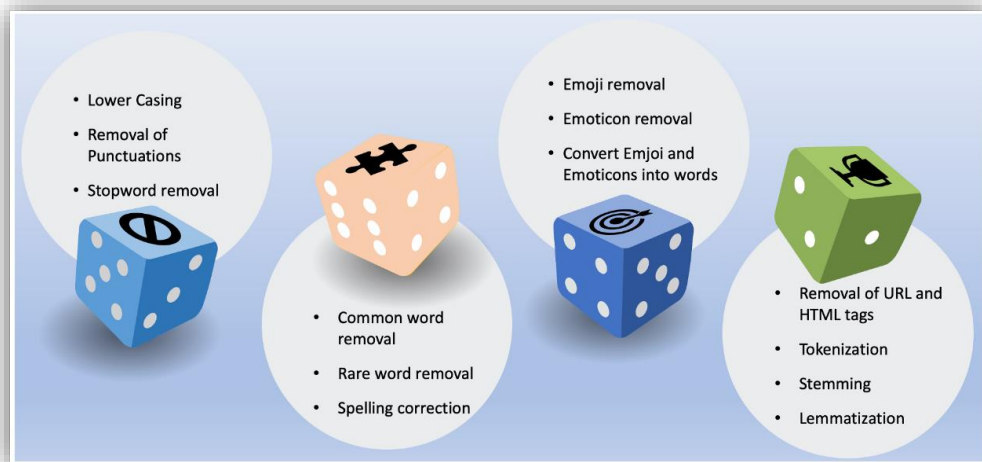
Berikut beberapa tipe thresholding:

- THRESH_BINARY.
- THRESH_BINARY_INV.
- THRESH_TRUNC.
- THRESH_TOZERO.
- THRESH_TOZERO_INV.

B. Pre-Processing Data Teks

Pra-pemrosesan teks adalah tugas penting dan langkah kritis dalam analisis teks dan pemrosesan bahasa alami (Natural Language Processing). Proses ini mengubah teks menjadi bentuk yang dapat diprediksi dan dianalisis sehingga algoritme pembelajaran mesin dapat bekerja lebih baik.

Ada berbagai cara untuk memproses teks sebelumnya. Berikut adalah beberapa pendekatan umum yang harus kita ketahui.



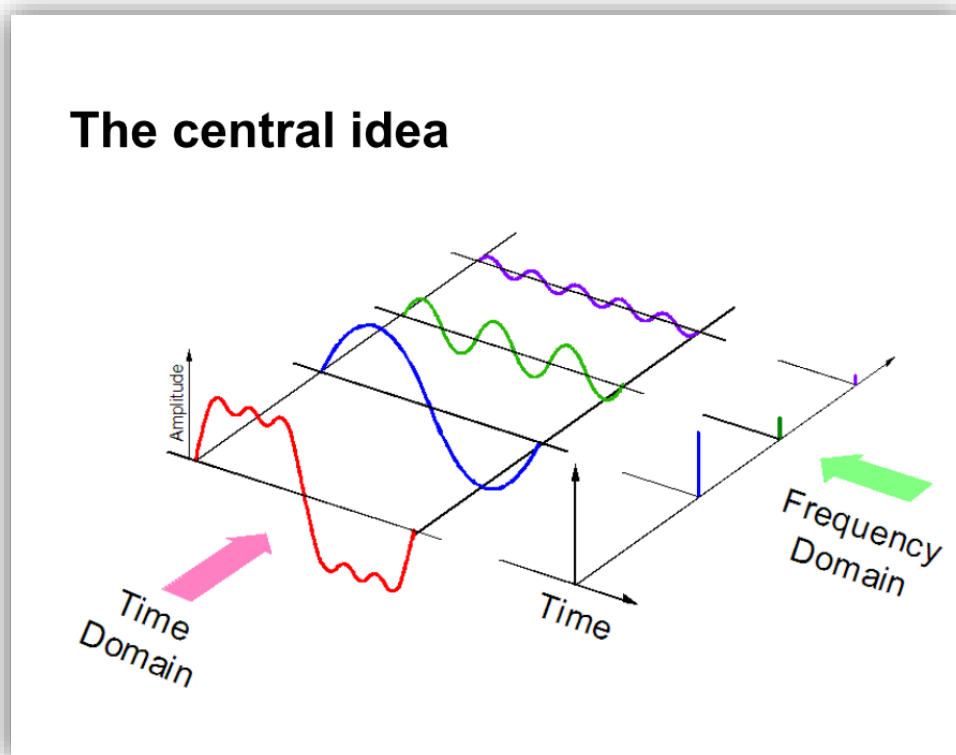
Dalam kasus ini kita menggunakan Tokenization, dimana Tokenisasi mengacu pada membagi teks menjadi urutan kata atau kalimat.

C. Pre-Processing Data Audio/Signal

Seperti semua format data tidak terstruktur, data audio memiliki beberapa langkah pra-pemrosesan yang harus diikuti sebelum disajikan untuk analisis. Langkah pertama adalah benar-benar memuat data ke dalam format yang dapat dimengerti mesin. Untuk ini, kita hanya mengambil nilai setelah setiap langkah waktu tertentu. Sebagai contoh dalam file audio 2 detik, kita mengekstrak nilai pada setengah detik. Ini disebut sampling data audio, dan kecepatan pengambilan sampelnya disebut sampling rate.

Cara lain untuk merepresentasikan data audio adalah dengan mengubahnya menjadi domain representasi data yang berbeda, yaitu domain frekuensi. Saat kita mengambil sampel data audio, kita membutuhkan lebih banyak titik data untuk mewakili seluruh data dan juga, laju pengambilan sampel harus setinggi mungkin.

Di sisi lain, jika kita merepresentasikan data audio dalam domain frekuensi, ruang komputasi yang dibutuhkan jauh lebih sedikit. Untuk mendapatkan intuisi, lihat gambar di bawah ini.



Source Code:

<https://github.com/trueazp/Data-Mining/blob/main/assignments/assignment06/Tugas06-Pertemuan7-PDM-H071191035.ipynb>