

FAKHRI FAUZAN / MidtransSEInternExam / 2018

Problem 1 [Topic : Text Mining]

Untuk mendapatkan similarities, hal yg perlu dilakukan adalah :

1. Perhatikan bahwa pada kasus ini berbentuk teks karena termasuk dalam Teks Mining.
2. Uraikan kalimat tersebut menjadi pecahan kata-kata (Split).
3. Kita perlu untuk menyiapkan kamus untuk kata-kata yang memiliki kesamaan makna. seperti "Jalan" dan "Jln" merupakan hal semakna. Karena mengandung makna yang sama maka bisa kita beri bobot dengan 1.
4. Setiap kata dihitung jumlah kemunculannya. Jika jumlah kemunculannya sama maka akan diberikan bobot 1, jika tidak akan diberi bobot 0.
5. Asumsikan atribut 1 dengan atribut lainnya tidak memiliki keterkaitan, agar bobot yang diperoleh pada setiap atribut dapat bernilai 0 atau 1 (Not Fuzzy System).
6. Contoh :

Jalan	Budi	Mulia	Gg.	F2	Nomor	29	D	RT.	04	RW	02	Jaksel	-
Jln	Budi	Mulia	Gg.	F2	Nomor	28	D	RT.	04	RW	02	Jaksel	IDN
1	1	1	1	1	1	0	1	1	1	1	1	1	0

$$\text{Sim} = 12/14 = 0.857 = 85.7\%$$