

Kuis (Latihan Soal)

Hitung nilai kemiripan antara Query dengan Dokumen
kemudian tentukan hasilnya, jika menggunakan Rumus

Cosinus Similarity:

- Dokumen1= Daun bewarna coklat
- Dokumen2= Bunga bewarna hijau
- Dokumen3= Rumput bewarna hijau
- Query= Daun hijau

Jawaban:

Merepresentasikan dokumen dan query dalam bentuk vektor dengan menghitung jumlah kemunculan setiap kata dalam dokumen atau query:

Dokumen1 = [1, 0, 1, 0] # Daun, bewarna, coklat, hijau

Dokumen2 = [0, 1, 0, 1] # Bunga, bewarna, hijau

Dokumen3 = [0, 0, 0, 1] # Rumput, hijau

Query = [1, 0, 0, 1] # Daun, hijau

Untuk Dokumen1: $(1 * 1) + (0 * 0) + (1 * 0) + (0 * 1) = 1 + 0 + 0 + 0 = 1$

Untuk Dokumen2: $(0 * 1) + (1 * 0) + (0 * 0) + (1 * 1) = 0 + 0 + 0 + 1 = 1$

Untuk Dokumen3: $(0 * 1) + (0 * 0) + (0 * 0) + (1 * 1) = 0 + 0 + 0 + 1 = 1$

Panjang vektor Query = $\sqrt{(1^2 + 0^2 + 0^2 + 1^2)} = \sqrt{2}$

Panjang vektor Dokumen1 = $\sqrt{(1^2 + 0^2 + 0^2 + 0^2)} = 1$

Panjang vektor Dokumen2 = $\sqrt{(0^2 + 1^2 + 0^2 + 1^2)} = \sqrt{2}$

Panjang vektor Dokumen3 = $\sqrt{(0^2 + 0^2 + 0^2 + 1^2)} = 1$

Cosine Similarity = (Produk Titik Query dan Dokumen) / (Panjang Vektor Query * Panjang Vektor Dokumen)

Untuk Dokumen1: Cosine Similarity1 = $1 / (\sqrt{2} * 1) = 1 / \sqrt{2}$

Untuk Dokumen2: $\text{Cosine Similarity}_2 = 1 / (\sqrt{2} * \sqrt{2}) = 1 / 2$

Untuk Dokumen3: $\text{Cosine Similarity}_3 = 1 / (1 * \sqrt{2}) = 1 / \sqrt{2}$

Jadi, hasil kemiripan antara Query dan Dokumen adalah sebagai berikut:

$\text{Cosine Similarity}_1 \approx 0.707$

$\text{Cosine Similarity}_2 = 0.5$

$\text{Cosine Similarity}_3 \approx 0.707$

Dengan demikian, Dokumen1 dan Dokumen3 memiliki kemiripan kosinus yang lebih tinggi dengan Query daripada Dokumen2.