Yohanes Dimas Pratama

A11.2021.13254 - A11.4705

***Model STKI dengan VSM***

Dokumen 1 = daun berwarna coklat

Dokumen 2 = bunga berwarna hijau

Dokumen 3 = rumput berwarna hijau

Query = daun hijau

Dokumen 1:

* Kata "daun" muncul 1 kali
* Kata "hijau" muncul 0 kali

Dokumen 2:

* Kata "daun" muncul 0 kali
* Kata "hijau" muncul 1 kali

Dokumen 3:

* Kata "daun" muncul 0 kali
* Kata "hijau" muncul 1 kali

Query:

* Kata "daun" muncul 1 kali
* Kata "hijau" muncul 1 kali

Vektor Dokumen 1: [1, 0]

Vektor Dokumen 2: [0, 1]

Vektor Dokumen 3: [0, 1]

Vektor Query: [1, 1]

Cosine Similarity Dokumen 1 = (1 + 0) / (√(1^2 + 1^2) \* √(1^2 + 0^2))

Cosine Similarity Dokumen 2 = (0 + 1) / (√(1^2 + 1^2) \* √(0^2 + 1^2))

Cosine Similarity Dokumen 3 = (0 + 1) / (√(1^2 + 1^2) \* √(0^2 + 1^2))

Cosine Similarity Dokumen 1 = 1 / (√2 \* 1) = 0.7071

Cosine Similarity Dokumen 2 = 1 / (√2 \* 1) = 0.7071

Cosine Similarity Dokumen 3 = 1 / (√2 \* 1) = 0.7071

Jadi, nilai kemiripan antara query "daun hijau" dan dokumen 1, dokumen 2, dan dokumen 3 adalah sekitar 0.7071 untuk masing-masing dokumen. Semakin tinggi nilai cosine similarity, semakin mirip dokumen dengan query.