

sumber baru	0	1	0	1	2	1,3	0	1,3	0	1,3
tanggul	0	1	1	1	3	1,1	0	1,1	1,1	1,1
panti	0	1	1	1	3	1,1	0	1,1	1,1	1,1
arjasa	0	1	1	1	3	1,1	0	1,1	1,1	1,1
pakusari	0	1	1	1	3	1,1	0	1,1	1,1	1,1
kalisat	0	1	1	0	2	1,3	0	1,3	1,3	0
ledokombo	0	1	1	0	2	1,3	0	1,3	1,3	0
sumberjambe	0	1	0	0	1	1,6	0	1,6	0	0
sukowono	0	1	1	1	3	1,1	0	1,1	1,1	1,1
jelbuk	0	1	1	1	3	1,1	0	1,1	1,1	1,1
kaliwates	0	1	1	1	3	1,1	0	1,1	1,1	1,1
sumber sari	0	1	1	1	3	1,1	0	1,1	1,1	1,1
patrang	0	1	1	1	3	1,1	0	1,1	1,1	1,1
kilojoule	0	1	1	1	3	1,1	0	1,1	1,1	1,1
kalori	0	1	1	1	3	1,1	0	1,1	1,1	1,1
lemak	0	1	1	1	3	1,1	0	1,1	1,1	1,1
protein	0	1	1	1	3	1,1	0	1,1	1,1	1,1
karbohidrat	0	1	1	1	3	1,1	0	1,1	1,1	1,1
sodium	0	1	0	1	2	1,3	0	1,3	0	1,3
kalium	0	1	1	1	3	1,1	0	1,1	1,1	1,1
kurang	0	2	0	1	2	1,3	0	2,6	0	1,3
risiko	0	1	0	0	1	1,6	0	1,6	0	0
senyawa	0	2	0	0	1	1,6	0	3,2	0	0
karotenoid	0	1	0	0	1	1,6	0	1,6	0	0
antioksidan	0	1	0	1	2	1,3	0	1,3	0	1,3
milik	0	1	0	0	1	1,6	0	1,6	0	0

mampu	0	1	0	0	1	1,6	0	1,6	0	0
bagai	0	1	0	1	2	1,3	0	1,3	0	1,3
anti	0	1	0	0	1	1,6	0	1,6	0	0
baik	0	4	0	1	2	1,3	0	5,2	0	1,3
derita	0	1	0	0	1	1,6	0	1,6	0	0
osteoarthritis	0	1	0	0	1	1,6	0	1,6	0	0
osteoporosis	0	1	0	0	1	1,6	0	1,6	0	0
turun	0	1	0	0	1	1,6	0	1,6	0	0
kolesterol	0	1	0	1	2	1,3	0	1,3	0	1,3
jahat	0	1	0	1	2	1,3	0	1,3	0	1,3
jaga	0	1	2	0	2	1,3	0	1,3	2,6	0
berat	0	1	1	0	2	1,3	0	1,3	1,3	0
badan	0	1	1	0	2	1,3	0	1,3	1,3	0
sehat	0	3	3	1	3	1,1	0	3,4	3,4	1,1
jantung	0	2	1	0	2	1,3	0	2,6	1,3	0
signifikan	0	1	0	0	1	1,6	0	1,6	0	0
janin	0	1	0	0	1	1,6	0	1,6	0	0
bagus	0	1	0	0	1	1,6	0	1,6	0	0
mata	0	1	0	0	1	1,6	0	1,6	0	0
stres	0	1	0	0	1	1,6	0	1,6	0	0
kingdom	0	1	1	1	3	1,1	0	1,1	1,1	1,1
plantae	0	1	1	1	3	1,1	0	1,1	1,1	1,1
divisi	0	1	1	1	3	1,1	0	1,1	1,1	1,1
magnoliophyta	0	1	0	0	1	1,6	0	1,6	0	0
kelas	0	1	1	1	3	1,1	0	1,1	1,1	1,1
magnoliopsida	0	1	1	1	3	1,1	0	1,1	1,1	1,1
ordo	0	1	1	1	3	1,1	0	1,1	1,1	1,1
laurales	0	1	0	0	1	1,6	0	1,6	0	0
famili	0	1	1	1	3	1,1	0	1,1	1,1	1,1
genus	0	2	2	2	3	1,1	0	2,2	2,2	2,2

species	0	1	1	1	3	1,1	0	1,1	1,1	1,1
anggur	0	0	7	0	1	1,6	0	0	11	0
hitam	0	0	2	0	1	1,6	0	0	3,2	0
vitis	0	0	4	0	1	1,6	0	0	6,4	0
vinifera	0	0	3	0	1	1,6	0	0	4,8	0
armenia	0	0	1	0	1	1,6	0	0	1,6	0
budidaya	0	0	2	0	1	1,6	0	0	3,2	0
kembang	0	0	1	0	1	1,6	0	0	1,6	0
timur	0	0	1	0	1	1,6	0	0	1,6	0
tengah	0	0	1	0	1	1,6	0	0	1,6	0
4000	0	0	1	0	1	1,6	0	0	1,6	0
bunga	0	0	1	0	1	1,6	0	0	1,6	0
majemuk	0	0	1	0	1	1,6	0	0	1,6	0
tumbuh	0	0	2	0	1	1,6	0	0	3,2	0
jenis	0	0	1	0	1	1,6	0	0	1,6	0
varietas	0	0	1	0	1	1,6	0	0	1,6	0
daerah	0	0	2	0	1	1,6	0	0	3,2	0
iklim	0	0	1	1	2	1,3	0	0	1,3	1,3
kering	0	0	1	0	1	1,6	0	0	1,6	0
datar	0	0	1	0	1	1,6	0	0	1,6	0
rendah	0	0	1	0	1	1,6	0	0	1,6	0
300	0	0	1	0	1	1,6	0	0	1,6	0
muka	0	0	1	0	1	1,6	0	0	1,6	0
laut	0	0	1	0	1	1,6	0	0	1,6	0
jombang	0	0	1	1	2	1,3	0	0	1,3	1,3
bangsalsari	0	0	1	0	1	1,6	0	0	1,6	0
atur	0	0	1	0	1	1,6	0	0	1,6	0
gula	0	0	1	0	1	1,6	0	0	1,6	0
darah	0	0	1	1	2	1,3	0	0	1,3	1,3
tingkat	0	0	3	1	2	1,3	0	0	3,9	1,3
fungsi	0	0	1	0	1	1,6	0	0	1,6	0

otak	0	0	1	0	1	1,6	0	0	1,6	0
lindung	0	0	1	0	1	1,6	0	0	1,6	0
rambut	0	0	1	0	1	1,6	0	0	1,6	0
daya	0	0	1	0	1	1,6	0	0	1,6	0
tahan	0	0	1	0	1	1,6	0	0	1,6	0
tubuh	0	0	1	0	1	1,6	0	0	1,6	0
tulang	0	0	1	0	1	1,6	0	0	1,6	0
kulit	0	0	1	1	2	1,3	0	0	1,3	1,3
spermatophyta	0	0	1	0	1	1,6	0	0	1,6	0
rhamnalses	0	0	1	0	1	1,6	0	0	1,6	0
famil	0	1	0	1	2	1,3	0	1,3	0	1,3
belimbing	0	0	0	6	1	1,6	0	0	0	9,6
averrhoa	0	0	0	3	1	1,6	0	0	0	4,8
carambola	0	0	0	2	1	1,6	0	0	0	3,2
pohon	0	0	0	1	1	1,6	0	0	0	1,6
kawasan	0	0	0	1	1	1,6	0	0	0	1,6
malaysia	0	0	0	1	1	1,6	0	0	0	1,6
negara	0	0	0	1	1	1,6	0	0	0	1,6
tropis	0	0	0	1	1	1,6	0	0	0	1,6
dunia	0	0	0	1	1	1,6	0	0	0	1,6
kencong	0	0	0	1	1	1,6	0	0	0	1,6
ajung	0	0	0	1	1	1,6	0	0	0	1,6
umbulsari	0	0	0	1	1	1,6	0	0	0	1,6
cerna	0	0	0	1	1	1,6	0	0	0	1,6
enzim	0	0	0	1	1	1,6	0	0	0	1,6
tekan	0	0	0	1	1	1,6	0	0	0	1,6
tinggi	0	0	0	1	1	1,6	0	0	0	1,6
ganggu	0	0	0	1	1	1,6	0	0	0	1,6

ginjal	0	0	0	1	1	1,6	0	0	0	1,6
kadar	0	0	0	1	1	1,6	0	0	0	1,6
nutrisi	0	0	0	1	1	1,6	0	0	0	1,6
makan	0	0	0	1	1	1,6	0	0	0	1,6
langsung	0	0	0	1	1	1,6	0	0	0	1,6
jerawat	0	0	0	1	1	1,6	0	0	0	1,6
magnoli	0	0	0	1	1	1,6	0	0	0	1,6
opsyta	0	0	0	1	1	1,6	0	0	0	1,6
geranial	0	0	0	1	1	1,6	0	0	0	1,6

Keterangan :

Q :Jumlah kemunculan *term* dalam query

D1 :Jumlah kemunculan *term* dalam dokumen 1

D2 :Jumlah kemunculan *term* dalam dokumen 2

D3 :Jumlah kemunculan *term* dalam dokumen 3

DF :Jumlah keseluruhan dokumen yang memuat *term*

IDF :bobot dari jumlah keseluruhan dokumen dibagi jumlah dokumen yang memuat *term*.

3.2.3 Perhitungan Cosine Similarity

Berikut langkah *retrieval* menggunakan data pada tabel perhitungan TF-IDF:

- a. Mengkalikan bobot antara bobot *query* dengan bobot *term* pada setiap dokumen

$$Q.D1 = (1*4)+(1*1)+(1*2) = 7$$

$$Q.D2 = (1*3)+(1*1)+(1*1) = 5$$

$$Q.D3 = (1*3)+(1*1)+(1*1) = 5$$

- b. Menghitung panjang *query* dengan mengakarkan jumlah kuadrat bobot *query*.

$$|Q| = \sqrt{1^2 + 1^2 + 1^2} = 1,732050808$$

- c. Menghitung panjang dokumen dengan mengakarkan jumlah dari bobot dokumen yang dikuadratkan pada setiap dokumen yang dikalikan dengan Q.

$$|D1| = \sqrt{(1^2 + 1^2 + 1^2) * (4^2 + 1^2 + \dots)} = 4,405094$$

$$|D2| = \sqrt{(1^2 + 1^2 + 1^2) * (3^2 + 1^2 + \dots)} = 4,388376$$

$$|D3| = \sqrt{(1^2 + 1^2 + 1^2) * (3^2 + 1^2 + \dots)} = 4,019487$$

Menghitung *cosine similarity* dengan membagi bobot Q.D_n dengan hasil perkalian antara panjang *query* (|Q|) dan panjang dokumen (|D_n|).

$$\text{Cos}(Q, D1) = \frac{7}{42,0174933} = 0,166597278$$

$$\text{Cos}(Q, D2) = \frac{5}{33,18139481} = 0,150686854$$

$$\text{Cos}(Q, D3) = \frac{5}{29,2252084} = 0,171085179$$

Ambang batas (threshold) adalah 0,16278977 yang di dapat dari jumlah total cosine di masing-masing dokumen di bagi banyaknya dokumen yaitu 3 Dokumen (D1, D2, D3). Sehingga tingkat similaritas dari skenario di atas yaitu D3, D1 dan D2.

3.3 Recall, Precision dan Accuracy

Keakuratan sistem dapat di evaluasi dengan cara perhitungan *Recall*, *Precision* dan *Accuracy* sebagai berikut :

Tabel 3.2 Perhitungan recall dan precision dengan threshold 0,09

	<i>Relevant</i>	<i>Not relevant</i>	Total
<i>Retrieve</i>	3	0	3
<i>not retrieve</i>	0	96	96
Total	3	96	99

Perhitungan *recall*, *precision* dan *accuracy* adalah sebagai berikut :

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{3}{3+0} = \frac{3}{3} \times 100 \\ &= \mathbf{100\%} \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{3}{3+0} = \frac{3}{3} \times 100 \\ &= \mathbf{100\%} \end{aligned}$$

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+FP+TN+FN} \\
 &= \frac{3+96}{3+0+96+0} = \frac{99}{99} \times 100 \\
 &= \mathbf{100\%}
 \end{aligned}$$

Dari contoh diatas dapat disimpulkan bahwa dengan *threshold* 0,1 sistem menghasilkan *recall* sebesar 100%, *precision* sebesar 100% dan *accuracy* 100%.

DAFTAR PUSTAKA

- Erawati, Indri, dan Yuni Yamasari. 2012. Aplikasi Ensiklopedia Negara Digital untuk Memotivasi Pengguna dalam Mengenal Negara di Dunia. *Jurnal Manajemen Informatika*.
- Hasugian, Jonner. 2006. Penelusuran Informasi Ilmiah Secara Online: Perlakuan Terhadap Seorang Pencari Informasi Sebagai Real User. *Jurnal Studi Perpustakaan dan Informasi*, 2(1), 1-13.
- Herdiawan. 2016. Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*.
- Komarayanti, Sawitri. 2017. *Ensiklopedia Buah-buahan Lokal Jember Berbasis Potensi Alam Jember*. p-ISSN 2527-7111; e-ISSN 2528-1615. <http://jurnal.unmuhjember.ac.id/index.php/BIOMA/article/download/591/470>.
- Komarayanti, Sawitri. 2018. Buah dan Sayur Lokal di Kabupaten Jember. ISBN : 978-602-6988-63-8.
- Melita, Ria dkk. 2018. Penerapan metode *term frequency inverse document frequency* (tf-idf) dan *cosine similarity* pada sistem temu kembali informasi untuk mengetahui *syarah hadits* berbasis web (studi kasus: *syarah umdatil ahkam*). *Jurnal Teknik Informatika*.
- Muharromah, Lailiyatul, Ulya Anisatur, Mudafiq Rihan Pratama. 2018. *Penilai Esai Otomatis Ujian Tengah Semester di SMK Asrama Pembina Masyarakat Jatimulyo Jember Menggunakan Metode Cosine Similarity*. Skripsi tidak diterbitkan. Jember: Program Studi Teknik Informatika.
- Salim, Agus, Wiwik Suharso, Hardian Oktavianto. 2018. *Pencarian Link Informasi Pada Aplikasi Ensiklopedia Buah dan Sayuran Lokal Dengan Metode Cosine Similarity*. Skripsi tidak diterbitkan. Jember: Program Studi Teknik Informatika.
- Sulistyo, Meiyanto Eko, Ristu Saptono, Adam Asshidiq. 2015. Penilaian Ujian Bertipe Essay Menggunakan Metode *Text Similarity*. *Jurnal Telematika*.
- Wahyudi, Dwi, Teguh Susyanto, Didik Nugroho. 2017. Implementasi dan Analisis Algoritma Stemming Nazief & Adriani dan Porter pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah Sinus* 15 (2), 49-56.