

Penerapan Cosine Similarity dan K-Nearest Neighbor (K-NN) pada Klasifikasi dan Pencarian Buku

Rut Samuel, Ripa Natan, Fitria, Umami Syafiqoh

Abstraksi— Penelitian ini bertujuan untuk membuat sebuah aplikasi Penerapan Cosine Similarity dan K-Nearest Neighbor (KNN) pada Klasifikasi dan Pencarian Buku. Dalam pencarian buku, user masih kesulitan dalam mencari buku referensi yang diinginkan.

Metode Cosine Similarity adalah mengukur kemiripan antara dua dokumen atau teks. Pada Cosine Similarity dokumen atau teks dianggap sebagai vector. Pada penelitian ini, Cosine Similarity digunakan untuk menghitung jumlah kata istilah yang muncul pada halaman-halaman yang diacu pada daftar indeks. Semakin banyak jumlah kata istilah yang muncul pada suatu halaman semakin tinggi nilai Cosine Similarity yang diperoleh.

Metode K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari contoh *query* yang baru diklasifikasi berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. Pada sistem ini menggunakan 100 data buku beserta sinopsis buku, yang masing-masing buku telah dibagi menurut klasifikasi buku dan diberikan bobot

Hasil klasifikasi dan pencarian buku dipengaruhi oleh *query* yang dimasukkan. Hasil klasifikasi dan pencarian buku membutuhkan waktu yang lama karena adanya proses *preprocessing*. Nilai akurasi pada pencarian dengan Cosine Similarity menghasilkan nilai rata-rata untuk *precision* 0.75 %, *recall* 0.6 % dan *f-measure* 0.66 % dengan menggunakan 3 sinopsis buku untuk menghitung nilai *precision*, *recall* dan *f-measure*. Untuk klasifikasi dengan K-Nearest Neighbor menghasilkan nilai rata-rata 70 % dengan menggunakan 10 data pembandingan.

Kata Kunci—Pencarian Buku, Klasifikasi, Cosine Similarity, K-Nearest Neighbor

I. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi tidak lepas dari meningkatnya kebutuhan manusia akan informasi yang cepat dan tepat. Teknologi informasi dapat dikatakan suatu teknologi yang digunakan untuk mengolah data, termasuk memproses, mendapatkan, menyusun, menyimpan, memanipulasi data dalam berbagai cara untuk menghasilkan informasi yang berkualitas.

Pada perpustakaan STMIK PPKIA TARAKANITA RAHMAWATI Tarakan, untuk klasifikasi buku dan pencarian buku, kebanyakan user sulit mendapatkan buku referensi yang sesuai dengan kebutuhan mereka di perpustakaan. Hal tersebut

dikarenakan kurangnya sistem yang dapat mempermudah proses pencarian buku yang sesuai dengan kebutuhan mereka. Oleh sebab itu diperlukan suatu sistem yang dapat mempermudah dalam pencarian buku. Salah satunya dengan pengklasifikasian buku sesuai dengan kategorinya sehingga proses pencariannya akan lebih akurat.

Klasifikasi buku merupakan proses pengelompokan buku sesuai dengan kategori yang dimilikinya. Sebuah buku dapat dikelompokkan ke dalam kategori tertentu berdasarkan kata/*query* pada judul buku tersebut. Kata atau kalimat yang terdapat di dalam sebuah buku memiliki makna tertentu dan dapat digunakan sebagai dasar untuk menentukan kategori buku tersebut. K-Nearest Neighbor (KNN) melakukan klasifikasi dengan berdasarkan ada penentuan *boundary* atau ruang lingkup secara lokal.

Dalam sistem ini, penulis menggunakan metode KNN untuk pengklasifikasian dan metode Cosine Similarity untuk pencarian buku. Jadi secara otomatis user dimudahkan dalam pencarian buku.

II. TINJAUAN PUSTAKA

Klasifikasi adalah proses pengelompokan, artinya mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama. Dalam kaitannya di dunia perpustakaan klasifikasi diartikan sebagai kegiatan pengelompokan bahan pustaka berdasarkan ciri-ciri yang sama [1].

Searching atau pencarian adalah proses pencarian data dari sekumpulan data yang sudah ada. Untuk data yang tidak ketemu biasanya ada prosedur tersendiri untuk menambah atau menyalin data yang belum ada tersebut [2].

Buku adalah beberapa helai kertas yang terjilid (berisi tulisan untuk dibaca atau halaman-halaman kosong untuk ditulis) Dari batasan di atas, dapat ditarik simpulan sementara bahwa buku adalah berisi bahan tertulis untuk memberikan pelajaran.

Buku adalah kumpulan kertas atau bahan lainnya yang dijilid menjadi satu pada salah satu ujungnya dan berisi tulisan atau gambar [3].

Cosine Similarity mengukur kemiripan antara dua dokumen atau teks. Pada Cosine Similarity dokumen atau teks dianggap sebagai vector. Untuk pencocokan teks, nilai dari vector A dan B adalah vektor term-frequency dari dokumen. Nilai cosine

similarity berada pada range 0-1. Pada penelitian ini, Cosine Similarity digunakan untuk menghitung jumlah kata istilah yang muncul pada halaman-halaman yang diacu pada daftar indeks. Semakin banyak jumlah kata istilah yang muncul pada suatu halaman semakin tinggi nilai Cosine Similarity yang diperoleh [4].

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Algoritma metode KNN sangatlah sederhana, bekerja dengan berdasarkan pada jarak terpendek dari sample uji ke sample latih untuk menentukan KNN nya. Setelah mengumpulkan KNN, kemudian diambil mayoritas dari KNN untuk dijadikan prediksi dari sample uji [5].

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Keterangan:

X_1 = Sampel Data

X_2 = data uji/testing

i = variabel data

d = jarak

p = dimensi data

Praproses teks adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Inputan awal pada proses ini adalah berupa dokumen.

Praprocessing teks pada penelitian ini terdiri dari beberapa tahapan, yaitu: proses pemecahan kalimat, proses tokenisasi, proses case folding, proses stopword removal, dan proses stemming [7]. Dalam penelitian ini stemming yang digunakan adalah stemming *Arifin dan Setiono*.

III. ANALISA DAN DESAIN SISTEM

A. Penerapan Metode

Perpustakaan merupakan salah satu gudang ilmu pengetahuan bagi mahasiswa dan dosen dalam lingkup kampus. Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan merupakan salah satu tempat bagi para mahasiswa maupun dosen untuk belajar dan menambah pengetahuannya melalui buku-buku yang ada. Untuk pencarian buku masih mengalami kesulitan dalam menemukan buku yang diinginkan.

Buku yang diklasifikasi memiliki keuntungan dalam penyusunan seperti buku yang sama akan diletakkan pada tempat yang berdekatan, memudahkan penelusuran terhadap bahan pustaka menurut subjek, memudahkan dalam penyusunan buku.

Dalam penerapan Cosine Similarity dan K-Nearest Neighbor yang harus dilakukan pertama kali adalah penentuan bobot untuk setiap dokumen. Langkah selanjutnya, adalah menghitung nilai bobot untuk setiap dokumen yang akan dicari. Langkah yang diambil admin pada sistem ini nantinya berawal dari user dalam mencari buku.

Pencarian buku dengan dicari berdasarkan kata kunci pada judul buku. Sistem menyiapkan data/sinopsis buku yang terdiri dari tahap case folding, tokenisasi, filtering, stop word removal, stemming dan tf-idf. Sebelum dilakukan perhitungan data yang

telah ada ditentukan kata kuncinya, kemudian tiap data ditentukan bobot masing-masing untuk setiap kata. Perhitungan bobot tiap kata dapat menggunakan pendekatan TF-IDF. Kemudian mencari nilai WDT dari hasil TF dan Log IDF, hasil wdt berlanjut ke perhitungan WD. Untuk panjang vector nilai pada kata kunci, D_1, D_2, D_3, D_n dihasilkan dari pangkat 2 dari perhitungan WDT.

Dari hasil perhitungan bobot dan panjang vector diperoleh nilai akhir dari data yang dicari, selanjutnya perbandingan dilakukan untuk menentukan total nilai tertinggi yang akan diambil untuk buku yang dicari.

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan pada jarak terpendek dari sample uji ke sample latih untuk menentukan KNN-nya. Setelah mengumpulkan KNN, kemudian diambil mayoritas dari KNN untuk dijadikan prediksi dari sample uji. Manfaat yang diperoleh dari klasifikasi dan pencarian buku ini adalah untuk memudahkan user dalam pencarian buku.

Jika dokumen-dokumen teks tersebut dikelompokkan (Classification) ke dalam dua kelas, C1 (Pemograman) dan C5 (Umum) maka kita dapat memperoleh hasil sebagai berikut:

C1 = akan beranggotakan D1, D2

C5 = akan beranggotakan D3

Pada k-NN, fase ini dinamakan fase manual atau training. Penulis memilih beberapa dokumen contoh (*sample*) dan mengelompokkannya secara manual ke dalam kelas-kelas yang telah didefinisikan. Pertanyaan, menggunakan K-NN tentukan klasifikasi dari query yang dimasukkan yaitu: logika algoritma,

Penyelesaian:

Langkah 1

Processing terhadap 3 dokumen yang terlibat D1, D2 dan D3 Dokumen yang akan diklasifikasi yaitu D4. Lakukan *tokenisasi, filtering, case folding, stop words removal dan stemming*. Hasilnya diperlihatkan pada tabel III. Tentukan bobot untuk setiap term dari 3 dokumen yang terlibat. Total dokumen ada 4. Dokumen yang telah terklasifikasi ada 3 dan yang akan diklasifikasikan (D4) sehingga total yang terlibat adalah 3.4.

Langkah 2

Hitung kemiripan vector dokumen D4 dengan setiap dokumen yang telah terklasifikasi (D1, D2, D3). Kemiripan antar dokumen dapat menggunakan cosine similarity. Hitung hasil perkalian scalar antar D4 dengan dokumen yang telah diklasifikasikan. Hasil perkalian dari setiap dokumen dengan D4 dijumlahkan (sesuai pembilang pada rumus diatas). Hitung panjang setiap dokumen termasuk D4 Caranya, kuadratkan bobot setiap term dalam setiap dokume, jumlahkan nilai kuadrat tersebut dan kemudian akarkan.

Hitung kemiripan D4 dengan D1, D2 dan D3

$\text{Cos}(D4, D1) = 0/(7.365*11.771) = 0$

$\text{Cos}(D4, D2) = 0/(7.365*7.407) = 0$

$\text{Cos}(D4, D3) = 1.242/(7.365*8.575) = 0.019$

TABLE I
 DATA AWAL

Dokumen	Kalimat	Kelas
1	Algoritma dan Pemrograman, Sebelum menyelami dunia pemrograman, ada pijakan yang harus Anda kuasai terlebih dahulu yaitu algoritma. Penguasaan algoritma sangat membantu Anda dalam memahami konsep-konsep penyusunan program sekaligus juga pemrograma. Buku ini hadir untuk menyajikan konsep-konsep algoritma dan pemnerapannya dalam proses penyusunan program, termasuk disini adalah konsep pemrograman objek. Pengantar Logika dan Algoritma, Buku ini akan sangat berguna bagi para pembaca umumnya dan mahasiswa khususnya sebagai bahan referensi menempuh perkuliahan karena disusun sedemikian rupa sehingga Agoritma sebagai dasar untuk menganalisis dan membentuk model matematis serta menyusun aspek proses logika dari suatu masalah tanpa mengacu pada suatu bahasa pemrograman tertentu mudah dipahami dan digunakan. Dengan demikian, pembaca atau mahasiswa dapat memahami konsep Logika serta dapat memanfaatkan Agoritma dalam menyelesaikan suatu kasus permasalahan secara sistematis dan logis.Buku ini disertai contoh aplikasi penerapan	C1
2	Logika dan Algoritma pada bahasa pemrograman Pascal sebab bahasa pada bahasa pada pemrograman tingkat tinggi (high level language) Pascal mempunyai Algoritma yang terstruktur dan pokok-pokok statementnya mewakili dari hampir semua pengertian tentang Logika dan Algoritma. Karena itu, mengerti tentang struktur perintah-perintah yang terdapat dalam bahasa pemrograman Pascal merupakan syarat untuk dapat memahami buku ini.Pokok-pokok yang dibahas dalam buku ini : Pengertian Dasar Logika dan Algoritma Konsep Tipe Data Diagram Alur atau Flowchart Struktur Rekursif Larik atau Array Metode Devide and Conquer Metode Greedy.	C1
3	Aplikasi logika fuzzy untuk pendukung keputusan.Aplikasi logika fuzzy sudah mulai dirasakan pada beberapa bidang. Salah satu aplikasi terpentingnya adalah untuk membantu manusia dalam melakukan pengambilan keputusan. Aplikasi logika fuzzy untuk pendukung keputusan ini semakin diperlukan tatkala semakin banyak kondisi yang menuntut Aplikasi logika fuzzy untuk pendukung keputusan. Aplikasi logika fuzzy untuk pendukung keputusan ini semakin diperlukan tatkala semakin banyak kondisi yang menuntut adanya keputusan yang tidak hanya bias dijawab dengan 'Ya' atau 'Tidak'. Buku berjudul Aplikasi Logika untuk Pendukung Keputusan ini mencoba untuk membantu para pembaca dalam memperluas wawasan tentang penggunaan logika fuzzy untuk pengambilan keputusan dengan uraian dalam bahasan beberapa metode langkah demi langkah.Buku ini membahas: Pendahuluan: berisi himpunan fuzzy dan operasi-operasinya.	C5

TABLE II
 DATA TESTING

Dokumen	Kalimat	Kelas
4	Kebutuhan informasi semakin mendesak sejalan dengan arus globalisasi yang terjadi di seluruh dunia. Keberhasilan informasi sangat tergantung pada sarana dan prasarananya. Pemakaian computer dalam kehidupan kita telah sangat meluas dan masyarakat, tidak terbatas di lingkungan kerja, tetapi juga dalam kehidupan sehari-hari. Dalam lingkungan perusahaan, computer adalah alat bantu yang mutlak diperlukan. Dalam penerapan system informasi yang berbasis compute menjadi kebutuhan mutlak untuk mendapatkan keunggulan kompetitif sehingga mendapat pririoritas yang tinggi.	C1

TABLE III
 HASIL CASE FOLDING

Dokumen	Kalimat
1	algoritma dan pemrograman sebelum menyelami dunia pemrograman ada pijakan yang harus anda kuasai terlebih dahulu yaitu algoritma penguasaan algoritma sangat membantu anda dalam memahami konsep konsep penyusunan program sekaligus juga pemrograma buku ini hadir untuk menyajikan ...
2	pengantar logika dan algoritma buku ini akan sangat berguna bagi para pembaca umumnya dan mahasiswa khususnya sebagai bahan referensi menempuh perkuliahan karena disusun ..
3	aplikasi logika fuzzy untuk pendukung keputusan aplikasi logika fuzzy sudah mulai dirasakan pada beberapa bidang salah satu aplikasi terpentingnya adalah untuk membantu manusia dalam melakukan pengambilan keputusan ...

No	Term	Tf					IDF Log (n/df)	Wdt = tf*idf			
		D4	D1	D2	D3	Df		D4	D1	D2	D3
1	analisa	1	0	0	0	1	1	1	0	0	0
2	sistem	2	0	0	0	1	1	2	0	0	0
3	inform asi	4	0	0	0	1	1	4	0	0	0
4	butuh	0	0	0	0	0	0	0	0	0	0
5	inform asi	3	0	0	0	1	1	3	0	0	0
6	desak	0	0	0	0	0	0	0	0	0	0
7	jalan	0	0	0	0	0	0	0	0	0	0
8	arus	1	0	0	0	1	1	1	0	0	0
9	global	0	0	0	0	0	0	0	0	0	0
10	dunia	1	0	0	0	1	1	1	0	0	0
11	hasil	0	0	0	0	0	0	0	0	0	0
12	inform asi	2	0	0	0	1	1	2	0	0	0
13	gantung	0	0	0	0	0	0	0	0	0	0
14	sarana	1	0	0	0	1	1	1	0	0	0
...
184	cocok	0	0	0	0	0	0	0	0	0	0
185	kecimpung	0	0	0	0	0	0	0	0	0	0

Gambar 3.1 Bobot Term

No	Term	Wd _i *w _{di}			Panjang Vektor			
		D1	D2	D3	D4 ²	D1 ²	D2 ²	D3 ²
1	analisa	0	0	0	1	0	0	0
2	sistem	0	0	0	4	0	0	0
3	informasi	0	0	0	16	0	0	0
4	butuh	0	0	0	0	0	0	0
5	informasi	0	0	0	9	0	0	0
6	desak	0	0	0	0	0	0	0
7	jalan	0	0	0	0	0	0	0
8	arus	0	0	0	1	0	0	0
9	global	0	0	0	0	0	0	0
10	dunia	0	0	0	1	0	0	0
11	hasil	0	0	0	0	0	0	0
12	informasi	0	0	0	4	0	0	0
13	gantung	0	0	0	0	0	0	0
...
184	cocok	0	0	0	0	0	0	0
185	kecimpung	0	0	0	0	0	0	0
		0.000	0.000	1.242	54.242	138.564	54.862	73.533
		Sqrt			7.365	11.771	7.407	8.575
		K-NN			0.000	0.000	0.020	

Gambar 3.2 Perhitungan Panjang Vector Untuk K-NN

TABLE IV
HASIL PERHITUNGAN

D1	D2	D3
0	0	0,019

Langkah 3: urutkan hasil perhitungan kemiripan, diperoleh:

TABLE V
HASIL PERANGKINGAN

1	2	3
D3	D1	D2

Ambil sebanyak k (k=3) yang akan dicari tingkat kemiripannya dengan query dan tentukan kelas dari query. Dokumen D4 terklasifikasi ke dalam kelas? Pilih kelas yang paling banyak kemunculannya! Apakah C1 atau C5. Ternyata, untuk k=3, C1 diwakili hanya oleh 2 dokumen yaitu D1 dan D2, sedangkan, C5, mewakili 1 dokumen, yaitu D3

Selanjutnya kita mencari query masuk pada klasifikasi yang mana apakah C1 atau C5. Dikarenakan nilai dari C5 lebih besar maka query masuk pada klasifikasi C5 yaitu Umum.

Cosine Similarity mengukur kemiripan antara dua dokumen atau teks. Pada Cosine Similarity dokumen atau teks dianggap sebagai vector. Untuk pencocokan teks, nilai dari vector A dan B adalah vektor term-frequency dari dokumen. Nilai cosine similarity berada pada range 0-1. Pada penelitian ini, Cosine Similarity digunakan untuk menghitung jumlah kata istilah yang muncul pada halaman-halaman yang diacu pada daftar indeks. Semakin banyak jumlah kata istilah yang muncul pada suatu halaman semakin tinggi nilai Cosine Similarity yang diperoleh.

Query yang dimasukkan dalam pencarian adalah “Logika Algoritma”

TABLE VI
DATA AWAL

Dokumen	Kalimat
1	algoritma dan pemrograman sebelum menyelami dunia pemrograman ada pijakan yang harus anda kuasai terlebih dahulu yaitu algoritma penguasaan algoritma sangat membantu anda dalam memahami konsep konsep penyusunan program sekaligus juga pemrograma buku ini hadir untuk menyajikan ...
2	pengantar logika dan algoritma buku ini akan sangat berguna bagi para pembaca umumnya dan mahasiswa khususnya sebagai bahan referensi menempuh perkuliahan karena disusun ..
3	aplikasi logika fuzzy untuk pendukung keputusan aplikasi logika fuzzy sudah mulai dirasakan pada beberapa bidang salah satu aplikasi terpentingnya adalah untuk membantu manusia dalam melakukan pengambilan keputusan ...

Data yang digunakan sudah telah melalui tahap case folding. Setelah tahap case folding, data diatas melalui tahap tokenisasi dan filtering seperti pada table VII

TABLE VII
TAHAP TOKENISASI DAN FILTERING

Data 1	Data 2	Data 3	Data 4	Data5
algoritma	Objek	Berguna	konsep	pengambilan
pemrograman	materi	pembaca	logika	keputusan
menyelami	materi	mahasiswa	memanfaatkan	aplikasi
aplikasi	dibahas	Bahan	algoritma	logika
...
proses penyusunan	objek	beberapa	salah	penggunaan
program	pengantar	bidang	aplikasi	logika
program	logika	salah	terpentingnya	
program	logika	salah	terpentingnya	
konsep	algoritma	mahasiswa	membantu	

Dari tahap tokenisasi dan filtering, selanjutnya ke tahap stemming

TABLE VIII
TAHAP STEMMING

Data 1	Data 2	Data 3	Data 4	Data5
algoritma	konsep	operasi	aspek	tingkat
program	program	matriks	proses	high
selam dunia	objek materi	algoritma struktur	logika mengacu	level language
...
integer	transportation	Bidang	transportation	problem
buku fuzzy	problem sisi	Batas Fuzzy	pandang metode	batas komputasi
program	buku	Buku	batas	bahas
cocok	fuzzy	Contoh	fuzzy	contoh

Selanjutnya dari tahap stemming, dilakukan tahap perhitungan bobot term dokumen

No	Term	Tf					IDF Log (n df)	Wdt = tf*idf			
		Q	D1	D2	D3	Df		Q	D1	D2	D3
1	logika	1	0	6	5	11	-0.564	0.000	-3.386	-2.821	
2	algoritma	1	4	7	0	11	-0.564	-0.564	-2.257	-3.950	
3	program	0	7	4	3	14	-0.669	0.000	-4.683	-2.676	
4	selam	0	1	0	0	1	0.477	0.000	0.477	0.000	
5	dunia	0	1	0	0	1	0.477	0.000	0.477	0.000	
6	pijak	0	1	0	0	1	0.477	0.000	0.477	0.000	
7	kuasa	0	1	0	0	1	0.477	0.000	0.477	0.000	
8	bantu	0	1	0	2	3	0.000	0.000	0.000	0.000	
9	paham	0	2	2	0	4	-0.125	0.000	-0.250	-0.250	
10	konsep	0	0	2	0	2	0.176	0.000	0.000	0.352	
11	susun	0	2	1	0	3	0.000	0.000	0.000	0.000	
12	buku	0	1	4	5	10	-0.523	0.000	-0.523	-2.092	
13	hadir	0	1	0	0	1	0.477	0.000	0.477	0.000	
14	saji	0	1	0	0	1	0.477	0.000	0.477	0.000	
15	terap	0	1	1	0	2	0.176	0.000	0.176	0.176	
16	proses	0	2	1	0	3	0.000	0.000	0.000	0.000	
...	
144	cocok	0	0	0	1	1	0.477	0.000	0.000	0.000	
145	kecimpung	0	0	0	1	1	0.477	0.000	0.000	0.477	

Gambar 3.3 Bobot Term

Hitung kemiripan vektor [dokumen] query Q dengan setiap dokumen yang ada. Kemiripan antar dokumen dapat menggunakan cosine similarity. Hitung hasil perkalian skala antara Q dan 3 dokumen lain. Hasilnya perkalian dari setiap dokumen dengan Q dijumlahkan. Hitung panjang setiap dokumen, termasuk Q. Caranya, kuadratkan bobot setiap term dalam setiap dokumen, jumlahkan nilai kuadrat dan terakhir akarkan.

No	Term	Tf					IDF Log (n df)	Wdt = tf*idf			
		Q	D1	D2	D3	Df		Q	D1	D2	D3
1	logika	1	0	6	5	11	-0.564	0.000	-3.386	-2.821	
2	algoritma	1	4	7	0	11	-0.564	-0.564	-2.257	-3.950	
3	program	0	7	4	3	14	-0.669	0.000	-4.683	-2.676	
4	selam	0	1	0	0	1	0.477	0.000	0.477	0.000	
5	dunia	0	1	0	0	1	0.477	0.000	0.477	0.000	
6	pijak	0	1	0	0	1	0.477	0.000	0.477	0.000	
7	kuasa	0	1	0	0	1	0.477	0.000	0.477	0.000	
8	bantu	0	1	0	2	3	0.000	0.000	0.000	0.000	
9	paham	0	2	2	0	4	-0.125	0.000	-0.250	-0.250	
10	konsep	0	0	2	0	2	0.176	0.000	0.000	0.352	
11	susun	0	2	1	0	3	0.000	0.000	0.000	0.000	
12	buku	0	1	4	5	10	-0.523	0.000	-0.523	-2.092	
13	hadir	0	1	0	0	1	0.477	0.000	0.477	0.000	
14	saji	0	1	0	0	1	0.477	0.000	0.477	0.000	
15	terap	0	1	1	0	2	0.176	0.000	0.176	0.176	
16	proses	0	2	1	0	3	0.000	0.000	0.000	0.000	
...	
144	cocok	0	0	0	1	1	0.477	0.000	0.000	0.000	
145	kecimpung	0	0	0	1	1	0.477	0.000	0.000	0.477	

Gambar 3.3 Perhitungan Panjang Vektor

Terapkan rumus cosine similarity. Hitung kemiripan Q dengan D1, D2 dan D3

$$\text{Cosine (Q, D1)} = \frac{1.274}{(0.977 * 5.678)} = 0.230$$

$$\text{Cosine (Q, D2)} = \frac{6.050}{(0.977 * 7.785)} = 0.795$$

$$\text{Cosine (Q, D3)} = \frac{3.184}{(0.977 * 8.828)} = 0.173$$

TABLE IX
 HASIL PERHITUNGAN

D1	D2	D3
0,230	0,795	0,173

Langkah 4: urutkan hasil perhitungan kemiripan, diperoleh:

TABLE X
 HASIL PERANGKINGAN

1	2	3
D2	D1	D3

Dokumen yang relevan dengan query “logika algoritma” yaitu D2

B. Uji Coba

Uji coba merupakan proses yang dilakukan untuk mencari perbedaan yang akan dijadikan perbandingan dari sistem yang telah dibuat. Pada uji coba kali ini penulis mencoba menguji hasil klasifikasi dan pencarian buku dengan perhitungan Precision, Recall dan F-Measure.

Setelah memperoleh klasifikasi dan pencarian buku, dilakukan perhitungan akurasi menggunakan precision, recall, dan f-measure. Precision adalah jumlah kelompok dokumen relevan dari total jumlah dokumen yang ditemukan oleh sistem atau menghitung jumlah dokumen relevan yang didapatkan sistem dibagi dengan jumlah dokumen dalam database, recall adalah menghitung jumlah dokumen relevan yang didapatkan sistem dibagi jumlah dokumen yang didapatkan sistem. F-Measure digunakan untuk mengukur kombinasi antara precision dan recall yaitu antara hasil dari presicion dan recall.

1. Uji coba untuk pencarian

Jumlah data yang digunakan 100 data, training data yang digunakan 95 dan data testing yang digunakan 5 data.

$$\text{Precision} = \frac{\text{jumlah dokumen relevan yang didapat sistem}}{\text{Jumlah dokumen yang didapatkan sistem}}$$

$$\text{Precision} = \frac{3}{4} = 0.75$$

$$\text{Recall} = \frac{\text{jumlah dokumen relevan yang didapatkan sistem}}{\text{Jumlah dokumen dalam database}}$$

$$\text{Recall} = \frac{3}{5} = 0.6$$

$$\text{F-measure} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}}$$

$$\text{F-measure} = \frac{2 \times (0.6 \times 0.75)}{0.6 + 0.75} = \frac{2 \times 0.45}{1.35} = \frac{0.9}{1.35} = 0.66$$

Dilakukan perhitungan untuk mencari nilai precision, recall, f-measure untuk sinopsis buku.

2. Uji coba untuk klasifikasi

Data yang akan diklasifikasikan ada 10 data. Dari 10 data yang diuji, ternyata hanya 7 data yang terklasifikasi dengan benar.

TABLE X
 DATA PEMBANDING

Sinopsis Buku	Kelas Manual	Kelas Sistem K-NN
Sinopsis 1	Statistika	Statistika
Sinopsis 2	Android	Android
Sinopsis 3	Robotic dan Elektronika	Robotic dan Elektronika
Sinopsis 4	Pemrograman Visual Basic	Pemrograman Visual Basic
Sinopsis 5	Desain Grafis	Desain Grafis
Sinopsis 6	Android	Statistika
Sinopsis 7	Robotic dan Elektronika	Robotic dan Elektronika
Sinopsis 8	Sistem Operasi	Umum
Sinopsis 9	Delphi	Robotic dan Elektronika
Sinopsis 10	Java	Java

$$\frac{7}{10} \times 100 = 70 \%$$

Dari hasil perbandingan data diatas, 10 data yang diinputkan ke sistem hanya 7 data yang terklasifikasi dengan benar. Jadi akurasi untuk K-Nearest Neighbor menghasilkan rata-rata 70 % dari data.

Pencarian dan klasifikasi buku pada penelitian ini menggunakan sinopsis yang diambil dari perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan dengan jumlah 100 judul dan sinopsis buku dengan memanfaatkan Cosine Similarity dan K-Nearest Neighbor (KNN). Hasil pencarian buku dipengaruhi oleh query yang dimasukkan. Hasil klasifikasi dan pencarian buku membutuhkan waktu yang lama karena adanya proses preprocessing.

Hasil uji coba untuk cosine similarity yang dilakukan menghasilkan rata-rata precision 0.75 %, recall 0.6 % dan f-measure 0.66 % digunakan 3 sinopsis buku untuk menghitung nilai precision, recall dan f-measure. Untuk klasifikasi buku, digunakan 10 data pembanding untuk melakukan uji cioba. Dan dari sinopsis buku yang diklasifikasikan hanya 7 data buku yang terklasifikasikan menurut sistem.

IV. KESIMPULAN

Dari pembahasan pada bab-bab sebelumnya maka dapat disimpulkan:

1. Dokumen dengan nilai >0 dan diurut dari nilai maksimal ke nilai minimal akan diambil sebagai data buku, kalimat yang diambil sebagai data buku merupakan kalimat yang mempresentasikan kesamaan pada kalimat dalam dokumen dengan query dan kesamaan antar kalimat dalam dokumen. Cosine Similarity digunakan untuk pencarian buku, dan K-Nearest Neighbor (KNN) untuk klasifikasi buku.
2. Hasil uji coba yang dilakukan menghasilkan rata-rata precision 0.75 %, recall 0.6 %, dan f-measure 0.66 % berdasarkan perbandingan pada jumlah data. Dan digunakan sebagai data dalam data buku sistem dan data buku manual, kemudian data hasil data buku sistem dibandingkan dengan data buku manual untuk memperoleh hasil akurasi.

3. Kelemahan karena semakin banyak dokumen yang akan dicari, semakin banyak waktu dalam pencarian dokumen

REFERENSI

- :
- [1] Gatot subrata S.Kom, Klasifikasi Bahan Pustaka diakses pada tanggal 4 Januari 2017 jam 09:15
 - [2] Mitsal Iqbal, Pengertian Searching dan Browsing, Bandung, 2013
 - [3] Mudzakir, Penulisan Buku Teks Bahasa Arab, Bandung, 2003
 - [4] Sherly Christina, Kinerja Cosine Similarity dan Semantic Similarity dalam Pengindentifikasian Releransi Nomor Halaman Pada Daftar Indeks Istilah, Yogyakarta, 2014
 - [5] Dr.Saed Sayad, K-Nearest Neighbor, University of Toronto, 2010
 - [6] B. Zamand, E.Winarko., Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia, IJCCS, 2011
 - [7] Jiwa Malem Marsya, Analisa dan Evaluasi Afiks Stemming untuk Bahasa Indonesia, FMIPA Universitas Syiah Kuala, 2011

Penulis I , Rut Samuel, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2017.

Penulis II , Ripa Natan, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2017.

Penulis III , Fitria, memperoleh gelar Magister Komputer (M.Kom), Sekolah Tinggi Teknik Surabaya. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati

Penulis IV , Ummi Syafiqoh, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2012. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati