

Sistem Rekomendasi Dosen Pembimbing berdasarkan Relevansi Topik Tugas Akhir menggunakan Metode Okapi BM25

Meilina Agustina*¹, Yufis Azhar², Nur Hayatin³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang
meylina.agus@gmail¹, yufis@umm.ac.id², noor.hayatin@umm.ac.id³

Abstrak

Sistem rekomendasi adalah sebuah perangkat lunak untuk memberikan rekomendasi kepada pengguna mengenai produk yang dapat digunakannya. Masalah administrasi di kantor jurusan Pendidikan Guru Sekolah Dasar Universitas Muhammadiyah Malang merupakan salah satu permasalahan yang selalu dihadapi oleh para staf TU dan part timer. Penggunaan sistem manual yang masih berjalan saat ini dinilai kurang efektif terhadap waktu, tempat, dan tenaga sehingga diperlukan adanya bantuan berupa sistem informasi. Pada perancangan sistem informasi ini akan menggunakan metode Okapi BM25 dimana metode ini merupakan fungsi peringkat yang digunakan oleh mesin pencari (search engine) untuk peringkat dokumen pencocokan sesuai relevansinya dengan permintaan pencarian yaitu berupa topik tugas akhir. BM25 memiliki fungsi yang sesuai dengan 3 prinsip pembobotan yang baik, yaitu memiliki inverse document frequency (idf), term frequency (tf), dan memiliki fungsi normalisasi dari panjang dokumen (document length normalization).

Kata Kunci: Okapi bm25, Sistem Rekomendasi, Text Mining

Abstract

The recommendation system is a software to provide recommendations to users about the products they can use. The administrative problem in the office of the Primary School Teacher Education department at the University of Muhammadiyah Malang is one of the problems faced by the Administration staff and part timers. The use of manual systems that are still running at this time is considered to be less effective against time, place, and energy, so that assistance in the form of information systems is needed. In designing this information system will use the Okapi BM25 method where this method is a ranking function used by search engines for matching document rankings according to their relevance to search queries, namely in the form of final assignment topics. BM25 has functions that are in accordance with the 3 principles of good weighting, which has an inverse document frequency (idf), term frequency (tf), and has a document length normalization function.

Keywords: Okapi bm25, Recommendation System, Text Mining

1. Pendahuluan

1.1 Latar Belakang

Teknologi informasi salah satu teknologi yang berkembang cepat saat ini. Penggunaan alat bantu komputer menjadi salah satu sarana penting untuk membantu menyelesaikan pekerjaan. Selain itu, komputer juga sebagai penunjang dalam sistem informasi yang dapat memberikan hasil lebih baik akurat untuk *output* ataupun input sebuah sistem, tentu bila sistem didalamnya telah berjalan dengan baik dan sesuai dengan yang diinginkan oleh *user*.

Jurusan PGSD (Pendidikan Guru Sekolah Dasar) adalah salah satu jurusan yang terdapat di Universitas Muhammadiyah Malang. Masalah administrasi di kantor jurusan merupakan salah satu permasalahan yang selalu dihadapi oleh para staf TU (Tata Usaha) dan *part timer*. Diantaranya adalah kesulitan pada saat pemilihan dosen pembimbing untuk mahasiswa tingkat akhir yang sedang mengajukan judul Tugas Akhir. Pada tahapannya mahasiswa diwajibkan untuk mengisi form pengajuan judul, kemudian mahasiswa juga diminta mencantumkan beberapa nama-nama dosen pembimbing yang nantinya akan dipilih oleh jurusan untuk menjadi dosen pembimbing mahasiswa tersebut. Penggunaan sistem manual yang masih berjalan saat inipun

dinilai kurang efektif terhadap waktu, tempat, dan tenaga sehingga diperlukan adanya bantuan berupa sistem informasi.

Ada beberapa penelitian yang sudah dilakukan sebelumnya. Sistem Repositori Tugas Akhir Mahasiswa dengan Fungsi Peringkat *Okapi BM25* oleh Ellysa Tjandra dan Monica Widiyasri. Dalam penelitian tersebut pencarian tugas akhir berdasarkan *query* yang diinput oleh pengguna menggunakan metode pencarian fungsi *Okapi BM25*. Dengan fungsi peringkat *Okapi BM25* maka hasil karya dapat ditampilkan dengan urutan peringkat sesuai relevansinya [1]. Begitu pula dengan Peringkat Dokumen Berbahasa Indonesia Berbasis Kata Benda dengan *BM25* oleh Rendy Rivaldi Pinandhita. Pada penelitian ini meliputi perhitungan koefisien dice antara hasil ringkasan manual dan hasil ringkasan sistem yang menggunakan *cosine similarity*, *Okapi BM25*, dan *content overlap*. Kinerja ringkasan dengan *Okapi BM25* lebih tinggi dari *cosine similarity* dan *content overlap*. Hasil terbaik dihasilkan oleh *Okapi BM25* pada percobaan tanpa menggunakan judul (*PageRank*). Lisa Noor Arida juga telah melakukan pengklasifikasian dosen pembimbing Tugas Akhir dengan menggunakan metode *Support Vector Machine (SVM)*, data yang digunakan berupa data Tugas Akhir Teknik Informatika UMM. Parameter data yang digunakan adalah judul tugas akhir, abstraksi, *keyword*. Penelitian ini difokuskan pada pengklasifikasian dosen pembimbing, sedangkan pada penelitian yang akan dikembangkan akan difokuskan pada pemberian rekomendasi dosen pembimbing Tugas Akhir. Adapun penelitian yang juga telah dilakukan oleh Dania Rahmanti yang melakukan penelitian dan perancangan sistem rekomendasi dosen pembimbing Tugas Akhir berdasarkan data keahlian dosen pembimbing. Data yang dibuat dari perancangan data matriks preferensi keahlian dosen pada jurusan Teknik Informatika Universitas Muhammadiyah Malang. Dengan data matriks tersebut dapat diketahui bahwa setiap dosen memiliki beberapa keahlian yang memiliki kemungkinan keahlian tersebut dimiliki oleh dosen lain [2].

Pada perancangan sistem informasi ini akan menggunakan metode *Okapi BM25* dimana metode ini merupakan fungsi peringkat yang digunakan oleh mesin pencari (search engine) untuk peringkat dokumen pencocokan sesuai relevansinya dengan permintaan pencarian yaitu berupa topik tugas akhir. Fungsi *Okapi BM25* merupakan fungsi yang memiliki tingkat keberhasilan terbaik sampai saat ini. *BM25* merupakan hasil dari percobaan beberapa variasi fungsi *Best Match* pada model peluang [3]. *BM25* memiliki fungsi yang sesuai dengan 3 prinsip pembobotan yang baik, yaitu memiliki *inverse document frequency (idf)*, *term frequency (tf)*, dan memiliki fungsi normalisasi dari panjang dokumen (*document length normalization*).

Rumusan Masalah

Dari uraian latar belakang tersebut adapun masalah-masalah yang akan diselesaikan dalam Tugas Akhir ini sebagai berikut:

1. Bagaimana merancang sebuah sistem rekomendasi dosen pembimbing menggunakan metode *Okapi BM25*?
2. Bagaimana menguji sistem yang telah dibuat?

1.2 Batasan Masalah

Untuk menghindari kesalahpahaman dan meluasnya pembahasan, maka permasalahannya hanya dibatasi dan difokuskan pada:

1. Data pada sistem ini hanya diambil dari *repository* resmi milik Universitas Muhammadiyah Malang yaitu pada divisi Fakultas Keguruan dan Ilmu Pendidikan jurusan PGSD Universitas Muhammadiyah Malang.
2. Sistem ini hanya fokus pada merekomendasikan dosen pembimbing berdasarkan relevansi topik Tugas Akhir.
3. Rekomendasi pada sistem ini hanya di fokuskan pada dosen pembimbing 1.
4. Rumus perhitungan yang digunakan adalah *Okapi BM25*.
5. Sistem dibangun sebatas simulasi dengan menggunakan *platform web*.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk melakukan rancang bangun Sistem Rekomendasi Dosen Pembimbing berdasarkan Relevansi Topik Tugas Akhir menggunakan Metode *Okapi BM25*.

1.4 Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini adalah sebagai berikut :

1. Studi Pustaka

Tahapan studi pustaka merupakan tahapan untuk mencari literatur – literatur berupa informasi dari artikel, jurnal, dan sumber lainnya yang berhubungan dengan implementasi *Okapi BM25*, pengolahan data *preprocessing*, serta perancangan sistem rekomendasi.

2. Analisis dan Desain Sistem

Membuat dan merancang sebuah Sistem Rekomendasi Dosen Pembimbing berdasarkan Relevansi Topik Tugas Akhir menggunakan Metode *Okapi BM25*.

3. Implementasi Pada Program

Pada tahap ini rancangan yang akan diimplementasikan pada Sistem Rekomendasi Dosen Pembimbing berdasarkan Relevansi Topik Tugas Akhir menggunakan Metode *Okapi BM25* berbasis Web.

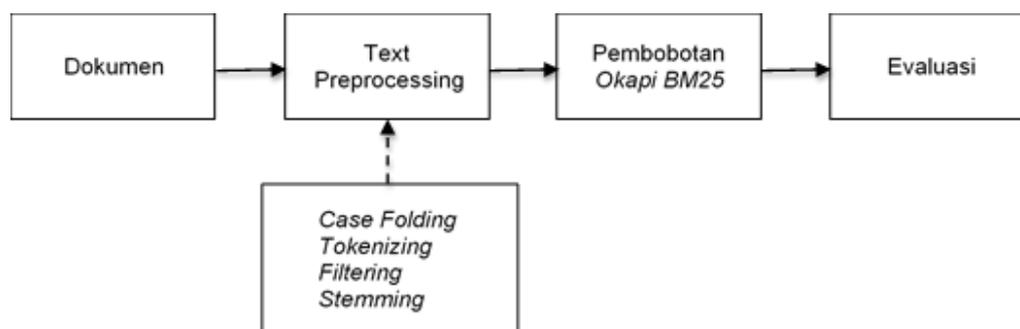
4. Evaluasi Program

Melakukan uji coba dan evaluasi terhadap sistem yang sudah dirancang. Proses ini diperlukan untuk memastikan bahwa tidak ada kesalahan dalam sistem dan untuk mengetahui kelebihan dan kekurangan dari sistem yang telah dibuat.

2. Tinjauan Pustaka

3.1 Sistem Rekomendasi

Sistem rekomendasi adalah sebuah perangkat lunak untuk memberikan rekomendasi kepada pengguna mengenai produk yang dapat digunakannya. Rekomendasi ini dibuat berdasarkan adanya personalisasi sehingga rekomendasi yang dihasilkan mungkin berbeda-beda bagi tiap *user*. Personalisasi ini dapat dihasilkan dari informasi *user*, yang berupa *rating* atau data transaksi *user* [4]. Gambar 1 berikut adalah alur sistem rekomendasi.



Gambar 1. Alur Sistem Rekomendasi

3.2 Text Mining

Text Mining merupakan salah satu aplikasi dari bidang data mining, yang khusus mengolah data dalam bentuk teks. Tujuan dari *text mining* adalah mencari informasi implisit dari data teks sehingga bisa digunakan oleh pengguna untuk mengambil keputusan [5].

3.3 Text Preprocessing

Tahap proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Sekumpulan karakter yang bersambungan (teks) harus dipecah-pecah menjadi unsur yang lebih berarti. *Preprocessing* pada tugas akhir ini terdiri dari beberapa tahapan yaitu: *case folding*, *tokenizing*, *filtering*, dan *stemming*.

3.3.1 Case Folding

Tahap *case folding* adalah tahap mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima.

3.3.2 Tokenizing

Tokenizing adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusunnya.

3.3.3 Filtering

Tahapan *filtering* adalah tahapan mengambil kata-kata penting dari dokumen. Bisa menggunakan algoritma *stoplist* (membuang kata yang tidak perlu) atau *wordlist* (menyimpan kata penting). *Stoplist* (*stopword*) adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopword* adalah “yang”, “dan”, “di”, “dari”, “kepada”, dan seterusnya.

3.3.4 Stemming

Stemming diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

3.4 Infomation Retrieval

Information Retrieval diartikan sebagai ilmu untuk menemukan material yang umumnya merupakan dokumen-dokumen, yang ditujukan untuk memenuhi kebutuhan informasi dari user. Dimana informasi tersebut tersebar didalam koleksi dokumen yang besar dan tidak terstruktur. Dengan adanya sistem temu kembali atau *information retrieval*, pengguna tidak lagi perlu melakukan pencarian dokumen secara manual, tetapi dapat dengan cepat mendapatkan informasi yang diinginkan melalui *query*.

Query merupakan token penting yang mewakili kebutuhan informasi dari seorang pengguna. *Query* diinputkan untuk diolah menggunakan metode yang diterapkan dalam sistem temu kembali melalui *search engine* [6].

3.5 Confusion Matrix

Confusion matrix adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasi dan jumlah data uji yang salah diklasifikasikan [7]. Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *presicion*, dan *recall*. *Accuracy* dalam klasifikasi adalah presentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [8]. Perhitungan akurasi dinyatakan dalam Persamaan 1 [7].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (1)$$

3.6 Okapi BM25

Di dalam *Information Retrieval*, *Okapi* merupakan sebuah algoritma yang digunakan oleh *search engine* untuk menentukan relevansi dokumen. Algoritma ini menggunakan *tf-idf* untuk melakukan prosesnya.

Tujuan akhir dari metode *Okapi BM25* sendiri adalah melakukan pembobotan dokumen berdasarkan *query* (atau dokumen yang menjadi patokan) yang dimasukkan dengan menghitung koefisien kemiripan (*Similarity Coefisien*) [6]. Rumus yang digunakan seperti pada Persamaan 2 [9].

$$RSV_{(d)} = \sum_{t \in Q} \log \left[\frac{N}{df_t} \right] \cdot \frac{(k_1 + 1) tf_{(td)}}{k_1 \left((1 - b) + b \left(\frac{L_d}{L_{ave}} \right) \right) + tf_{(td)}} \quad (2)$$

Keterangan:

$RSV_{(d)}$: *Retrieval Status Value* untuk suatu dokumen.

$\sum_{t \in Q} \log \left[\frac{N}{df_t} \right]$: disebut juga nilai *idf term t*.

N : jumlah dokumen dalam korpus

df_t : jumlah dokumen dalam korpus yang mengandung *term* pada *query q*.

$tf_{(td)}$: frekuensi *term t* yang muncul pada dokumen *d*.

L_d : panjang dokumen *d*.

L_{ave} : rata-rata panjang dokumen secara keseluruhan.

k_1 : konstanta frekuensi *term*, nilai $1,2 \leq k_1 \leq 2$.

b : konstanta panjang dokumen, nilai $0 \leq b \leq 1$

3. Analisa dan Perancangan Sistem

3.1 Analisa

3.1.1 Analisa Data

Dalam sistem rekomendasi ini diperlukan adanya beberapa data berupa data Tugas Akhir mahasiswa diantaranya yaitu seperti nama mahasiswa, judul Tugas Akhir, abstraksi, nama dosen pembimbing 1, dan nama dosen pembimbing 2. Data yang digunakan didalam penelitian ini bersumber dari *repository* resmi milik Universitas Muhammadiyah Malang.

3.1.2 Analisa Sistem

Pemilihan dosen pembimbing Tugas Akhir masih dilakukan mahasiswa secara manual yang dinilai kurang efektif terhadap waktu, tempat, dan tenaga menyebabkan kurang optimalnya pengambilan keputusan pemilihan dosen pembimbing. Pada tahapannya mahasiswa diwajibkan untuk mengisi form pengajuan judul Tugas Akhir, kemudian mahasiswa juga diminta mencantumkan beberapa nama-nama dosen pembimbing yang nantinya akan dipilih oleh jurusan untuk menjadi dosen pembimbing mahasiswa tersebut. Mahasiswa memilih dosen pembimbing hanya berdasarkan keahlian dosen yang mereka ketahui. Kurang adanya data preferensi dan keahlian dosen sebagai acuan mahasiswa untuk menentukan pilihan dosen pembimbing dapat menimbulkan berbagai masalah diantaranya mahasiswa ragu dalam memilih dosen perekomendasi Tugas Akhir, ide mahasiswa sulit dikembangkan, serta tingkat resiko tinggi pemilihan dosen pembimbing yang spesifikasinya kurang tepat terhadap ide maupun judul Tugas Akhir yang diajukan oleh mahasiswa.

3.2 Perancangan Fitur

Perancangan fitur dilakukan untuk melakukan perancangan awal sistem, bagaimana sistem mengolah data sampai siap untuk digunakan.

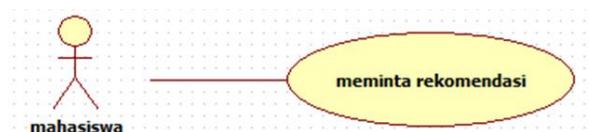
3.2.1 Preprocessing Data

Preprocessing data dilakukan pada data uji. Data uji dalam penelitian ini merupakan data baru yang diinputkan oleh mahasiswa berupa data abstraksi singkat atau ide Tugas Akhir. Tujuannya adalah untuk menyaring kata-kata menjadi lebih terstruktur dan menormalisasi kata-kata tidak baku menjadi kata dasar dan siap digunakan untuk proses pada tahap selanjutnya [10].

3.2.2 Perancangan Sistem

3.2.2.1 Usecase

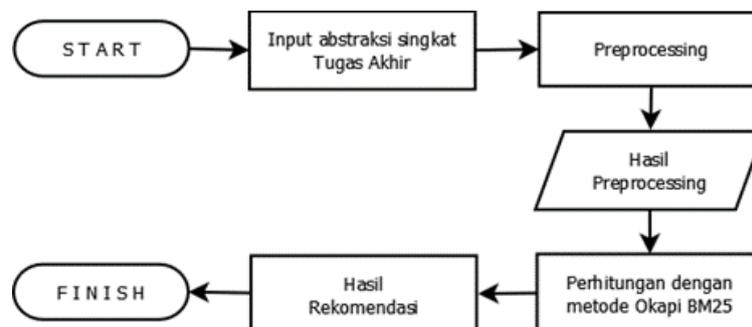
Aktor pada *usecase* pada Gambar 2 adalah Mahasiswa. Terdiri dari 1 (satu) *usecase* yaitu meminta rekomendasi.



Gambar 2. Usecase Diagram Sistem Rekomendasi

3.2.2.2 Flowchart

Flowchart secara keseluruhan Tugas Akhir ini ditunjukkan pada Gambar 3.



Gambar 3. Flowchart Alur Sistem

3.2.3 Perancangan Pengujian

Pengujian menggunakan metode *confusion matrix*, dimana evaluasi menggunakan *confusion matrix* akan menghasilkan nilai akurasi (*accuracy*). Akurasi dalam klasifikasi adalah presentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi.

True positive (TP) merupakan jumlah *record positive* dalam data set yang diklasifikasikan *positive*. *True negative* (TN) merupakan jumlah *record negative* dalam data set yang diklasifikasikan *negative*. *False Positive* (FP) merupakan jumlah *record negative* dalam data set yang diklasifikasikan *positive*. *False Negative* (FN) merupakan jumlah *record positive* dalam data set yang diklasifikasikan *negative*. Tabel 1 berikut adalah tabel persamaan dari model *confusion matrix*:

Tabel 1. Confusion Matrix

		Kelas Prediksi	
		<i>Relevant</i> (1)	<i>Non Relevant</i> (0)
Kelas Sebenarnya	<i>Retrieved</i> (1)	True Positives (<i>TP</i>)	False Positives (<i>FP</i>)
	<i>Not Retrieved</i> (0)	False Negatives (<i>FN</i>)	True Negatives (<i>TN</i>)

Nilai akurasi (*accuracy*) menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar.

4. Implementasi dan Hasil

4.1 Implementasi Proses

Tahap implementasi ini menjelaskan bagaimana langkah awal dari proses data mentah yang kemudian diolah keseluruhan secara manual dengan menjadi sesuatu yang diharapkan penulis.

4.2 Proses Data Mentah

Gambar 4 merupakan data mentah dari topik tugas akhir yang digunakan untuk merekomendasikan dosen pembimbing.

	A	B	C	D
1	No	Penulis	Judul	Abstraksi
1		RENIKA ARISINTA	PENGEMBANGAN MODEL GROUP INVESTIGATION DENGAN PERMAINAN "AKU SEORANG DETEKTIF" DALAM MATERI GAYA DAN GERAK PADA KELAS IV SD MUHAMMADIYAH 1 MALANG	Hasil dari observasi dan wawancara ditemukan pada tanggal 10 Desember 2016 yaitu menghasilkan (1) guru kurang berinteraksi dengan siswa dan metode pembelajaran yang digunakan masih metode ceramah namun pernah menggunakan model Group Investigation; (2) Siswa kurang aktif dalam berkomunikasi terkait pembelajaran; dan (3) Guru masih berpacu pada pada Buku Siswa dan Buku Guru. Sehingga perlunya pengembangan pada model pembelajaran yang sudah pernah digunakan yaitu mengembangkan model pembelajaran Group Investigation dengan permainan "Aku Seorang Detektif" dengan kriteria valid, efektif dan menarik. Penelitian ini bertujuan untuk menghasilkan model Group Investigation (GI) dengan permainan "Aku Seorang Detektif" dalam materi gaya dan gerak pada kelas IV SD Muhammadiyah 1 yang layak dengan kriteria valid, efektif, dan menarik. Jenis penelitian yang dilakukan adalah Research and Development (RnD). Dalam mengembangkan model GI, digunakan rancangan ADDIE Model karena model ini disusun secara terprogram dengan urutan-urutan kegiatan yang sistematis dalam upaya pemecahan masalah belajar yang berkaitan dengan sumber belajar yang sesuai dengan kebutuhan dan karakteristik pembelajar. Pada ADDIE Model juga terdapat evaluasi pada setiap langkahnya. Sehingga meminimalisir kesalahan pada produk ADDIE Model terdiri atas lima langkah, yaitu: (1) analisis, (2) perancangan, (3) pengembangan, (4) penerapan, (5) evaluasi. Hasil penelitian menunjukkan (1) Pengembangan model pembelajaran GI dengan permainan "aku seorang detektif" dalam materi
2		KIKI AFRILIANI	ANALISIS PEMANFAATAN SUMBER BELAJAR BERBASIS LINGKUNGAN DALAM PEMBELAJARAN TEMATIK TEMA 8 (EKOSISTEM) KELAS V SDN MOJOLANGU 5 MALANG	Mengenalikan lingkungan sekitar sebagai suatu potensi untuk belajar merupakan salah satu wujud menanamkan kepada siswa sikap cinta lingkungan, dan melestarikan alam, meskipun saat ini dihadapkan suatu kehidupan yang serba modern. Oleh karena itu pemanfaatan sumber belajar berbasis lingkungan perlu diterapkan didalam pembelajaran, sebab didalam pembelajaran tematik siswa dihadapkan pada pembelajaran yang kontekstual dan membawa siswa untuk lebih mengenal lingkungan. Penelitian ini bertujuan untuk: "Mendeskripsikan (perencanaan, pelaksanaan dan evaluasi) yang dilakukan oleh guru dalam pemanfaatan sumber belajar berbasis lingkungan pada pembelajaran tematik tema 8 (Ekosistem) kelas V SDN Mojolangu 5 Malang". Penelitian ini menggunakan penelitian kualitatif dengan jenis penelitian deskriptif. Data dalam penelitian ini adalah pemanfaatan sumber belajar dalam pembelajaran tematik tema 8 (Ekosistem) di kelas V. Sumber data yaitu guru kelas V dan siswa-siswi kelas V SDN Mojolangu 5 Malang. Pengumpulan data menggunakan wawancara, observasi, dan dokumentasi. Kegiatan analisis data yang digunakan yaitu pengumpulan data, reduksi data, penyajian data, dan verifikasi/kesimpulan. Hasil penelitian Pada SD Muhammadiyah 06 Lawang Kabupaten Malang terdapat guru yang menerapkan keterampilan mengajar. Keterampilan mengajar terdiri dari 8 keterampilan yaitu keterampilan
3				

Gambar 4. Data Mentah

4.2.1 Proses Preprocessing

Pada tahap ini data mentah melalui beberapa proses preprocessing yaitu, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

4.2.2 Proses Case Folding

Pada Gambar 5 merupakan proses *case folding* yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil.

7						
8	bahan	ajar	yang	digunakan	di	lapangan
9	penggunaan	media	sangat	membantu	dalam	proses
10	penelitian	ini	bertujuan	untuk	mendeskripsikan	penggunaan
11	sistem	pembelajaran	di	sekolah	dasar	menuntut
12	perubahan	kurikulum	2006	menjadi	kurikulum	2013
13						

Gambar 5. Proses Case Folding

4.2.3 Proses Tokenizing

Gambar 6 adalah proses dari *tokenizing* dimana pemotongan *string* input berdasarkan tiap kata yang menyusunnya.

	A	B	C	D	E	F
1	Bahan	ajar	yang	digunakan	di	lapangan
2	Penggunaan	media	sangat	membantu	dalam	proses
3	Penelitian	ini	bertujuan	untuk	mendeskripsikan	penggunaan
4	Sistem	pembelajaran	di	Sekolah	Dasar	menuntut
5	Perubahan	kurikulum	2006	menjadi	kurikulum	2013
6						

Gambar 6. Proses Tokenizing

4.2.4 Proses Filtering dan Stemming

Pada Gambar 7 merupakan hasil dari proses *filtering* yaitu, tahapan mengambil kata-kata penting dari dokumen.

	A	B	C	D	E
1	aktif	absah	ahli	ajar	ajar
2	alpha	ajar	ahli	ajar	ajar
3	analisis	ajar	ahli	ajar	ajar
4	angket	ajar	ahli	ajar	ajar
5	belakang	ajar	ahli	ajar	ajar
6	bukti	ajar	ahli	ajar	ajar
7	data	ajar	ahli	ajar	ajar
8	deskriptif	ajar	ahli	ajar	ajar
9	deskriptif	ajar	ahli	ajar	ajar
10	dominan	ajar	ahli	ajar	ajar

Gambar 7. Hasil Filtering

Kemudian dari hasil filtering akan dilanjutkan ke tahapan *stemming* seperti pada Gambar 8, yaitu proses memperkecil jumlah indeks yang berbeda dari suatu dokumen dan mengelompokkan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena imbuhan.

	A	B	C	D	E
1	aktif	absah	ahli	ajar	ajar
2	alpha	ajar	ahli	ajar	ajar
3	analisis	ajar	ahli	ajar	ajar
4	angket	ajar	ahli	ajar	ajar
5	belakang	ajar	ahli	ajar	ajar
6	bukti	ajar	ahli	ajar	ajar
7	data	ajar	ahli	ajar	ajar
8	deskriptif	ajar	ahli	ajar	ajar
9	deskriptif	ajar	ahli	ajar	ajar
10	dominan	ajar	ahli	ajar	ajar

Gambar 8. Hasil Stemming

Pada Gambar 9 menunjukkan hasil dari preprocessing data training yang telah diberi label secara manual. Dari hasil tersebut dapat dilihat bobot *transformation TF-IDF* ada masing – masing kata. Dimana prosesnya menghitung kehadiran atau ketidakhadiran sebuah kata didalam dokumen.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1																									
2	No	Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21	D22	D23
3	1	absah			1			1																1	
4	2	abstrak																							3
5	3	acak	1																						
6	4	action																							
7	5	acu																							
8	6	acuan																							
9	7	adaptasi																							
10	8	addie																		1					
11	9	adil	1																						
12	10	adobe																							
13	11	afektif									1														
14	12	ahli			2			12		6	3		6	5					6			5		3	
15	13	ajak														2									
16	14	ajar	17	5	7		15	16	24	3	19	13	5	10	9	12	11	1	16	3	1		6	6	
17	15	akademik																				1			
18	16	akan																							
19	17	akrab													1										
20	18	aktif	2			1		1				1													2
21	19	aktivitas																							
22	20	akui																							
23	21	alamat														1	1								

Gambar 9. Hasil Preprocessing Data Training

4.3 Implementasi dan Hasil Eksperimen Algoritma Okapi BM25

4.3.1 Implementasi Pembobotan Okapi BM25

Pada Gambar 10 merupakan proses implementasi secara umum algoritma Okapi BM25 dalam Microsoft Excel. Data didalam operator excel menggunakan hasil preprocessing data training sebanyak 80 data dokumen dan menggunakan data uji sebanyak 20 data query. Kemudian data training diklasifikasikan menggunakan pembobotan Okapi BM25 berdasarkan query.

	FE	FF	FG	FH	FI	FJ	FK	FL	FM	FN	FO	FP	FQ	FR
1						Okapi-BM25								
2	D77	D78	D79	D80	Avgdl	D1	D2	D3	D4	D5	D6	D7	D8	D9
3	129	131	54	158	113,3625	0	0	0,597144787	0	0	0,588872532	0	0	0
4						0	0	0	0	0	0	0	0	0
5						0	1,103189446	0	0	0	0	0	0	0
6						0	0	0	0	0	0	0	0	0
7						0	0	0	0	0	0	0	0	0
8						0	0	0	0	0	0	0	0	0
9						0	0	0	0	0	0	0	0	0
10						0	0	0	0	0	0	0	0	0
11						0	0,928687391	0	0	0	0	0	0	0
12						0	0	0	0	0	0	0	0	0
13						0	0	0	0	0	0	0	0	0,690068666
14						0	0	0	0,43282082	0	0	0,615307454	0	0,589557384
15						0	0	0	0	0	0	0	0	0
16						0,125301714	0	0,106089841	0,113109421	0	0,122300605	0,120802421	0,126367526	0,097343016
17						0	0	0	0	0	0	0	0	0
18						0	0	0	0	0	0	0	0	0
19						0	0	0	0	0	0	0	0	0
20						0,461243206	0	0	0	0,335434586	0	0	0,312733422	0
21						0	0	0	0	0	0	0	0	0
22						0	0	0	0	0	0	0	0	0
23						0	0	0	0	0	0	0	0	0

Gambar 10. Hasil Pembobotan Okapi BM25

4.3.2 Pengujian Akurasi

Pengujian akurasi dilakukan untuk mengetahui seberapa akurat metode dalam memberikan hasil rekomendasi. Skenario pengujiannya menggunakan 20 data abstrak yang berbeda sebagai data uji.

Perhitungan akurasi:

$$\begin{aligned}fx &= 13 & fn &= 20 \\ \text{Akurasi} &= \frac{fx}{fn} \times 100\% \\ &= \frac{13}{20} \times 100\% \\ &= 65\%\end{aligned}$$

Keterangan:

fx = total hasil rekomendasi sesuai

fn = total keseluruhan hasil rekomendasi

Dari pengujian dengan menggunakan 20 data uji yang berupa data abstrak Tugas Akhir yang diujikan pada sistem, hasil akurasi kesesuaian sistem dalam memberikan rekomendasi sebesar 65 %.

Referensi

- [1] Tjandra, E., & Widasri, M. (2016). Sistem Repositori Tugas Akhir Mahasiswa dengan Fungsi Peringkat Okapi BM25. *Journal of Information Systems Engineering and Business Intelligence*, 2(2), 88. <https://doi.org/10.20473/jisebi.2.2.88-94>
- [2] Pinandhita, R. R. (2013). *Peringkat Dokumen Berbahasa Indonesia Berbasis Kata Benda Dengan BM25*.
- [3] Frinta, K., & Adikara, P. P. (2019). Pencarian Berita Berbahasa Indonesia Menggunakan Metode BM25. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2589–2595.
- [4] Hapsari, Y. G., Wibowo, A. T., Informatika, F., Telkom, U., Informatika, F., & Telkom, U. (2015). *Analisis Dan Implementasi Sistem Rekomendasi Menggunakan Most-Frequent Item Dan Association Rule Technique Analysis and Implementation Recommender System Using Most-Frequent Item and Association Rule Technique*. 2(3), 7757–7764.
- [5] Lumbanraja, F. R. (2013). *Sistem Pencarian Data Teks dengan Menggunakan Metode Klasifikasi Rocchio (Studi Kasus : Dokumen Teks Skripsi)*.
- [6] Sebastian, D., Rachmat, A., & Sudiarto Raharjo, W. (2011). Implementasi Algoritma Okapi BM25 dan K-Means untuk Mencari Relevansi Artikel pada Beberapa Situs Berita. *Seminar Teknik Informatika Dan Sistem Informasi*, 1–14.
- [7] Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 5–10.
- [8] Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Seminar Nasional Sistem Informasi Indonesia*, (November), 2–3.
- [9] Bintana, R. R. (2012). *Penerapan Model Okapi Bm25 Pada Sistem Temu Kembali Informasi*. 273–279.
- [10] Prahudaya, T. Y., & Harjoko, A. (2017). A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia. *Jurnal Teknosains*, 6(2), 113. <https://doi.org/10.22146/teknosains.26972>

