
Pemanfaatan *Text Mining* Dalam Penentuan Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi (Studi Kasus: Universitas Bina Darma)

Depta Rahma Mulia¹, Yesi Novaria Kunang²

Program Studi Sistem Informasi, Fakultas Sains Teknologi
Universitas Bina Darma

email : 181410288@student.binadarma.ac.id¹, yesinovariakunang@binadarma.ac.id²,

Jl. A. Yani No. 3, Palembang 30624, Indonesia

Abstract

The final assignment is the end of a student's journey in college where this final assignment is one of the determinants of whether or not students graduate. To be able to make the final assignment well, students get a supervisor who must be in accordance with the topic of the student's final assignment. Students studying at Bina Darma University will of course pass what is called a final project regardless of the major they take. In order to get supervisors who are in accordance with the topic of the student's final assignment, a basic system is created to recommend supervisors. This research was conducted using the BM25 method which is a method by means of ranking where the higher the BM25 score the lecturer's name will be at the top and it is most likely that the lecturer can be selected for students who input the topic of the final project. With raw data obtained from lecturer research data and student final assignment data several years ago, totaling 1473 data. From this data, it is then processed with text preprocessing and information retrieval. And then weighted with TF-IDF and finally, ranking is carried out by applying the program code from the BM25 score so that the results of the lecturer's recommendations are obtained.

Kata kunci: *BM25, recommendation, text mining*

Abstrak

Tugas akhir adalah akhir dari perjalanan mahasiswa dalam masa kuliah yang mana tugas akhir ini menjadi salah satu penentu lulus tidaknya mahasiswa. Untuk dapat membuat tugas akhir dengan baik mahasiswa mendapat dosen pembimbing yang harus sesuai dengan topik tugas akhir dari mahasiswa tersebut. Mahasiswa yang berkuliah di Universitas Bina Darma tentunya akan melewati yang namanya tugas akhir dimanapun jurusan yang mereka ambil. Untuk mendapatkan dosen pembimbing yang memang sesuai dengan topik tugas akhir mahasiswa, dibuatlah suatu sistem dasar untuk merekomendasikan dosen pembimbing. Penelitian ini dilakukan dengan menggunakan metode BM25 yang merupakan sebuah metode dengan cara perankingan yang mana semakin tinggi skor BM25 tersebut nama dosen akan berada diatas dan kemungkinan besar dosen tersebut dapat terpilih bagi mahasiswa yang meng-input topik tugas akhir. Dengan data mentah yang didapat dari data penelitian dosen dan data tugas akhir mahasiswa beberapa tahun lalu yang berjumlah 1473 data. Dari data tersebut kemudian diproses dengan *text preprocessing* dan *information retrieval*. Setelah itu, dilakukan pembobotan dengan TF-IDF dan terakhir dilakukan perankingan dengan menerapkan kode program dari BM25 *score* sehingga terdapatlah hasil rekomendasi dosen..

Kata kunci: *BM25, rekomendasi, teks mining*

1. PENDAHULUAN

Teknologi berpengaruh besar pada kemajuan lembaga pendidikan dan dibutuhkan sistem informasi manajemen yang optimal untuk memberikan pelayanan terbaik bagi civitas akademik. Untuk mendapatkan lulusan terintegritas, langkahnya termasuk membuat tugas akhir. Skripsi adalah hasil penelitian mahasiswa dengan bimbingan dosen pembimbing. Pembimbing sangat penting untuk menyusun skripsi berkualitas bagi mahasiswa di universitas negeri atau swasta.

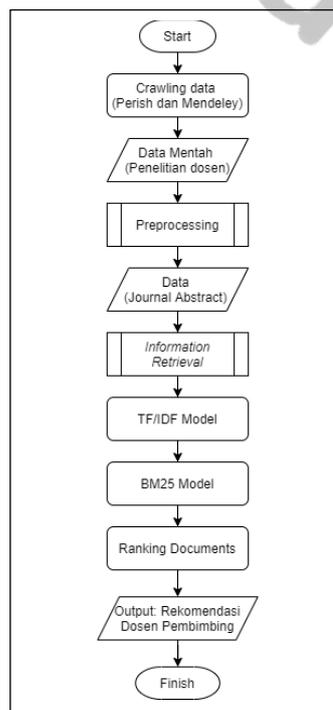
Program Studi Sistem Informasi Univeristas Bina Darma Palembang memiliki mata kuliah konsentrasi Data Science dan penelitian yang terkait. Mahasiswa perlu mengajukan proposal dan dipilih dosen yang cocok agar dapat melakukan bimbingan pra-proposal. Sistem merekomendasikan dosen dengan text mining yang sesuai dengan topik yang diajukan dengan cara mengevaluasi kesesuaian judul yang diajukan dengan penelitian dosen melalui pemeringkatan kecocokan.

Sebelum penelitian ini dilakukan, penulis mengacu kepada beberapa penelitian terdahulu meliputi Sistem Rekomendasi Penentuan Dosen Pembimbing Tugas Akhir dengan Menggunakan Algoritma Rabin-Karp (Salam, Wicaksana and Hastuti, 2015), Sistem Rekomendasi Pemilihan Dosen Pembimbing Menggunakan Metode *Weighted Product* (Yulianti, Sutomo and Perdana, 2020), *Text Mining Classification* sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi (Efendi and Mustakim, 2017).

Dengan adanya penelitian sebelumnya yang menggunakan beberapa metode seperti Naïve Bayes, K-nn, SVM, dan mungkin masih banyak yang lainnya. Maka dari itu, penelitian yang dibuat ini mencoba untuk menggunakan metode BM25 yang merupakan sistem pemeringkatan yang digunakan untuk mengurutkan *output* kecocokan terhadap dokumen-dokumen.

2. METODOLOGI PENELITIAN

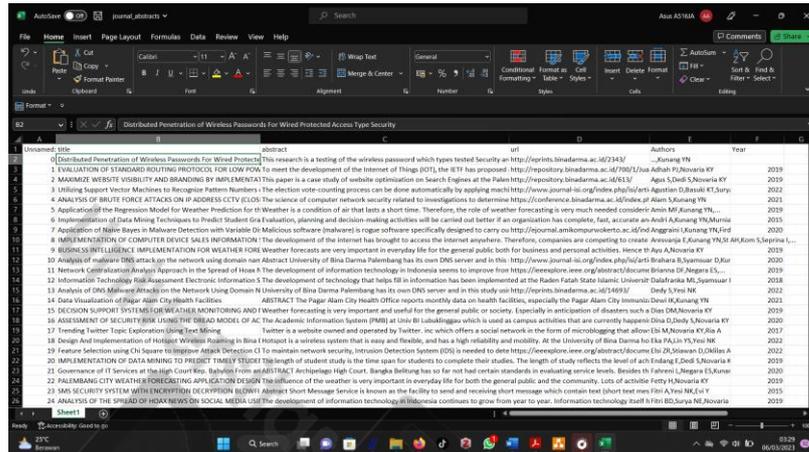
Metode penelitian yang dilakukan menggunakan metode eksperimen dengan skema penelitian sebagai berikut:



Gambar 1. Skema Penelitian

2.1 Pengumpulan Data (Crawling Data)

Pengumpulan data dilakukan dengan menggunakan teknik *crawling* pada *Publish/Perish* berupa data teks. Dengan data dosen sebanyak 33 dosen jurusan Sistem Informasi dan jumlah penelitian sebanyak 1473 data.



Gambar 2. Data Mentah

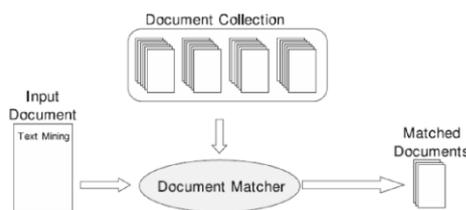
2.2 Text Preprocessing

Preprocessing sendiri merupakan tahapan mengubah teks asli sebagai masukan dan menerapkan beberapa rutinitas dasar untuk mengubah atau menghilangkan unsur tekstual yang tidak berguna dalam pengolahan lebih lanjut (Najjichah, Syukur and Subagyo, 2019). Tahapan ini dilakukan dengan mengisi data dari proses *crawling*. Pada penelitian ini, *preprocessing* dilakukan dengan 4 tahapan yang meliputi

1. *Case folding*, proses mengubah semua huruf dalam dokumen menjadi bentuk yang standar.
2. *Remove stopwords*, penghapusan kata yang dapat digunakan pada saat merepresentasikan dokumen. *Stopword* sendiri berisi kumpulan kata yang dianggap tidak deskriptif dan bisa dihilangkan menggunakan pendekatan *bag-of-words* (Arisetiawan, 2019)
3. *Tokenizing*, memcah aliran konten tekstual menjadi kata, istilah, simbol, atau elemen makna lainnya (Vijayarani and Janani, 2016).
4. *Stemming*, memperoleh informasi dengan sistem pengambilan seperti mesin pencari dan analitik teks (Mustaqim, Umam and Muslim, 2020)

2.3 Information Retrieval

Information Retrieval dilakukan dengan menyediakan *query* teks pencarian yang diinginkan terhadap kumpulan dokumen untuk menemukan dokumen yang relevan.



Gambar 3. Proses Information Retrieval

Tujuan dari pencarian ini adalah untuk mencari dokumen yang sesuai dengan kebutuhan pengguna. Adapun penjelasan mengenai tahapan pada *information retrieval*.

1. *Feature statistic*, menghitung panjang suatu dataset.
2. *Indexing framework*, data akan dibuat terstruktur dengan tujuan mempermudah dalam menemukan informasi secara cepat pada saat diminta.
3. *TF/IDF*, algoritma yang berasal dari bidang *information retrieval*, namun saat ini semaik banyak digunakan dalam perbandingan dokumen (Andayani and Ryansyah, 2017). Algoritma digunakan untuk menentukan bobot dari suatu kata pada suatu dokumen dapat dilihat pada persamaan berikut:

$$tfidf_{t,d} = tf_{t,d} \cdot idf_t \quad (1)$$

Keterangan:

$tfidf_{i,d}$ = bobot term

$tf_{i,d}$ = *term frequency* kata pada dokumen

idf_i = *inverse document frequency* kata

4. BM25 Model, menentukan seberapa tinggi tingkat kemiripan antara dataset yang ada dengan *query* yang nantinya menjadi tolak ukur pencarian data yang diinginkan. BM25 menghasilkan keefektifan dan ketepatan yang tinggi untuk menghitung kesamaan antara *query* dan dokumen (Pardede, Husada and Riansyah, 2021). Dalam perhitungan tingkat kesamaan dokumen, BM25 menggunakan 3 faktor antara lain TF, IDF, rata-rata panjang dokumen (Pranata, 2019), dapat dilihat pada persamaan 2 dibawah ini:

$$BM25 = \sum_{i=1}^{|q|} idf_{(q_i)} \cdot \frac{tf_{(q_1d)} \cdot (k_1 + 1)}{tf_{(q_1d)} + k_1 \cdot (1 - b + b \cdot \frac{|d|}{dl_{ave}})} \quad (2)$$

Keterangan:

$idf_{(q_i)}$: nilai *invers document frequency* pada *term query*.

$tf_{(q_1,d)}$: jumlah frekuensi *term query* i pada koleksi dokumen.

K_1 : $1,2 \leq k_1 \leq 2,0$

B : $0,5 \leq b \leq 0,8$

dl_{ave} : rata-rata Panjang semua koleksi dokumen

$|d|$: Panjang koleksi dokumen

5. *Query formulation*, memasukkan *query* yang digunakan sebagai tolak ukur pencarian dokumen.
6. *Rangking documents*, melakukan pemeringkatan dokumen yang sesuai dengan *query* yang telah dimasukkan sebelumnya dengan ketentuan berupa adanya *precision*, *recall*, dan *F1-score*.
7. *Evaluation metrics*, uji model dari program yang telah dibuat sebelumnya dengan penentuan apakah sistem tersebut dapat berjalan semestinya.

2.4 Penerapan Algoritma BM25

Penerapan algoritma ini membutuhkan *output* dari data yang telah diolah dari proses sebelumnya yaitu hasil dari proses *preprocessing* dan hasil dari pembobotan kata dengan *Tf-idf*.

2.5 Hasil Rekomendasi

Hasil rekomendasi akan didapat pada saat seluruh tahapan sebelumnya selesai dan telah dievaluasi dengan hasil akhir mengenai tingkat *accuracy*, *precision*, dan *confussion matrix* dari pengkategorian ekcockan setelah dilakukan pelabelan.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Proses

Dalam pengimplementasian proses ini *tools* yang digunakan adalah *google colaboratory*. *Tools* ini adalah sebuah editor yang berbentuk web yang mampu berjalan di komputer, ada beberapa hal yang bisa dilakukan google colab sendiri seperti menulis *code python*, visualisasi, dan juga sebagai *markdown* editor.

Dataset yang berjumlah 1473 penelitian dosen dimasukkan ke google colab untuk dibaca datanya. Untuk tampilan datanya dibuat batasan hingga 1000 data. Dengan tampilan berikut

Unnamed: 0	0	title	abstract	url	Authors	Year
0	0	Distributed Penetration of Wireless Passwords ...	This research is a testing of the wireless pas...	http://eprints.binadarma.ac.id/2343/	..., Kunang YN	NaN
1	1	EVALUATION OF STANDARD ROUTING PROTOCOL FOR LO...	To meet the development of the Internet of Thi...	http://repository.binadarma.ac.id/700/1/juan%2...	Adhah PJ, Novaria KY	2019.0
2	2	MAXIMIZE WEBSITE VISIBILITY AND BRANDING BY IM...	This paper is a case study of website optimiza...	http://repository.binadarma.ac.id/613/	Agus S, Dedi S, Novaria KY	2019.0
3	3	Utilizing Support Vector Machines to Recognize...	The election vote-counting process can be done...	http://www.journal-isi.org/index.php/isi/artic...	Agustian D, Basuki KT, Surya NE, ...	2022.0
4	4	ANALYSIS OF BRUTE FORCE ATTACKS ON IP ADDRESS ...	The science of computer network security relat...	https://conference.binadarma.ac.id/index.php/B...	Alam S, Kunang YN	2021.0
...
1469	1469	SALES EXECUTIVE INFORMATION SYSTEM IN COMPUTER...	RKU Computer expects quality improvement in sa...	NaN	Fatmasari F	2022.0
1470	1470	ANALYSIS OF INTENSITY BEHAVIOR OF USING E-LEAR...	Ethan Frome ANALYSIS OF BEHAVIORAL INTENSITY O...	http://eprints.binadarma.ac.id/14918/1/judu%2...	Fatmasari F, Purnamasari SD	2022.0
1471	1471	Implementation of the e-learning system using ...	This study aims to determine behavioral intent...	NaN	Fatmasari F	2011.0
1472	1472	Application of the Technology Acceptance Model...	This research was conducted to determine the c...	NaN	Fatmasari F, Ariandi RM	2014.0
1473	1473	Application of the Technology Acceptance Model...	This research was conducted to determine the c...	https://www.neliti.com/publications/224998/pen...	Fatmasari F, Ariandi M	2014.0

Gambar 4. Output Proses Pembacaan Data

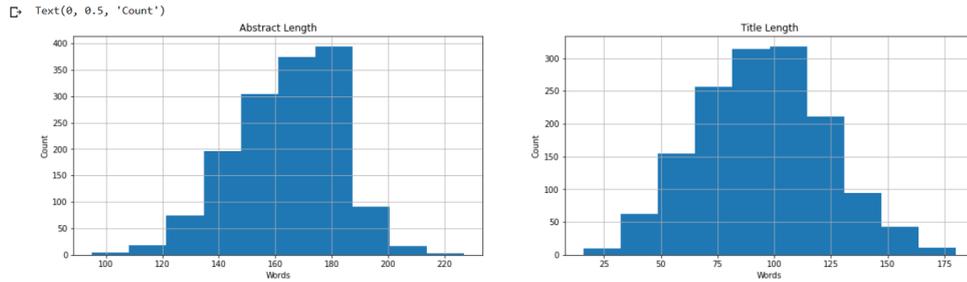
3.2 Implementasi Information Retrieval dan Algoritma BM25

1. *Feature statistic* digunakan untuk menghitung panjang data yang ada didalam dataset dengan ketentuan pada bagian abstrak dan *title* dihitung panjangnya.

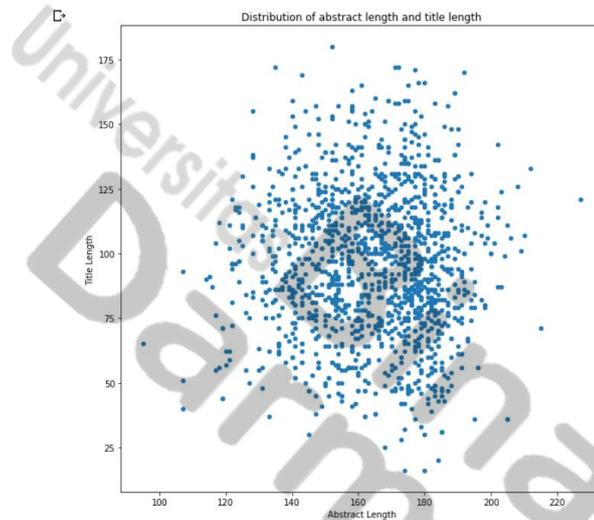
	Unnamed: 0	Year	abstract_length	title_length
count	1474.000000	1401.000000	1474.000000	1474.000000
mean	736.500000	2018.899358	164.014247	94.692673
std	425.651461	3.075805	18.550873	27.877165
min	0.000000	2005.000000	95.000000	16.000000
25%	368.250000	2017.000000	151.000000	76.000000
50%	736.500000	2020.000000	166.000000	94.000000
75%	1104.750000	2021.000000	178.000000	114.000000
max	1473.000000	2023.000000	227.000000	180.000000

Gambar 5. Hasil dari Feature Statistic

Selanjutnya dilakukan pemvisualisasian data dengan memberikan gambar berupa grafik bar dengan ketentuan 2 grafik yaitu grafik *abstract length* dan *title length*. Selain kedua grafik tersebut, adapun grafik *distribution of abstract length and title length*.



Gambar 6. Grafik Panjang Abstrak dan Title



Gambar 7. Grafik of abstract length and title length

2. *Indexing Framework* bertujuan membuat dokumen agar mudah untuk dirangkingkan oleh *search engine*. Dataset yang ada akan kembali dibaca dan dijabarkan menjadi seperti pada tahapan *loading dataset* hanya saja pada tahapan ini akan langsung dibuat perangkingan.
3. TF-IDF Model, dilakukan pemberian skor pada dokumen yang telah ditentukan vektornya dengan ketentuan gambar yang muncul sebanyak 10 suku kata, hasil lebih jelas dapat dilihat pada gambar dibawah ini.

	2010	2014	85	abilities	able	abstract	academic	access	accesses	accident	...	wireless	wlans	wold	women	works	world	wpa	writing	written	year
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.823079	0.0	0.0	...	0.572989	0.0	0.0	0.0	0.000000	0.0	1.25813	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	1.119845	0.0	0.000000	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.955709	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0

Gambar 8. Hasil Model TF-IDF

4. BM25 Model, pada penelitian ini diberikan ketentuan berupa 2 model BM25 yaitu *default model* dan *specific model*. Yang bertujuan untuk memberikan pilihan model mana yang lebih bagus.

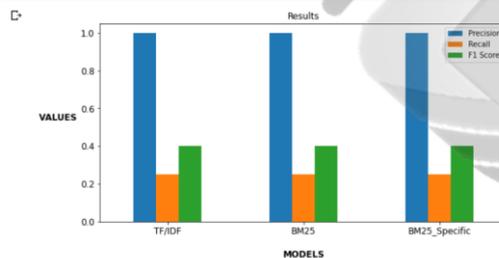
Gambar 9. Hasil Pembobotan Model BM25 Default dan Specific

5. *Query Formulation*, untuk memberikan sebuah *query* atau kata kunci yang fungsinya memperkecil pencarian artikel penelitian. Dengan contoh *query*-nya *wireless* dan *internet*.
6. *Ranking Documents*, dilakukan untuk pemberiang ranking pada saat evaluasi metrik nantinya. Terdapat *precision* k dan *F1-score* yang dipakai pada proses selanjutnya untuk perbandingan antara model TF-IDF, BM25, BM25 *specific*.
7. *Evaluation Metrics*, mengevaluasi program dan fungsi yang telah dibuat dengan memanfaatkan *query* dan juga perhitungan pada tahap *ranking documents*. Dan didapat hasil index model dan juga hasil visualisasinya sebagai berikut:

TF/IDF retrieved query "internet" with Precision @ 7 = 0.8, Recall = 0.8, and F1-score = 0.8000000000000002
 TF/IDF retrieved query "wireless" with Precision @ 7 = 1.0, Recall = 0.25, and F1-score = 0.4
 Default BM25 retrieved query "internet" with Precision @ 7 = 0.8, Recall = 0.8, and F1-score = 0.8000000000000002
 Default BM25 retrieved query "wireless" with Precision @ 7 = 1.0, Recall = 0.25, and F1-score = 0.4
 Specific BM25 retrieved query "internet" with Precision @ 7 = 0.8, Recall = 0.8, and F1-score = 0.8000000000000002
 Specific BM25 retrieved query "wireless" with Precision @ 7 = 1.0, Recall = 0.25, and F1-score = 0.4

	Precision	Recall	F1 Score
Model			
TF/IDF	1.0	0.25	0.4
BM25	1.0	0.25	0.4
BM25_Specific	1.0	0.25	0.4

Gambar 10. Tabel Perbandingan Hasil Query



Gambar 11. Hasil visualisasi

Setelah itu akan digunakan fungsi dari `retrieve_ranking` untuk melihat dokumen mana yang relevan dengan pencarian *wireless*.

```
[65] print(*doc_scores[:10], sep = "\n")
(39, 3.865324771098428)
(47, 3.865324771098428)
(18, 2.8101167533867004)
(41, 2.730203170283565)
(80, 2.730203170283565)
(0, 2.654709043048244)
(1, 0.0)
(2, 0.0)
(3, 0.0)
(4, 0.0)
```

Gambar 12. Skor BM25 Pencarian Queries Wireless

Pada gambar tersebut didapatkan skor BM25 beserta dengan nomor dokumen yang relevan dengan pencarian *queries*. Untuk hasil dokumen yang didapat bisa dilihat pada gambar dibawah ini.

```
SCORE BM25 RANKING id 10 termirip
[39, 47, 18, 41, 80, 0, 1, 2, 3, 4]

SCORE BM25 RANKING Title 10 termirip
39 RADIUS-Based Wireless Authentication System Se...
47 RADIUS-Based Wireless Authentication System Se...
18 Design And Implementation of Hotspot Wireless ...
41 Development of a Centralized Academic Hotspot ...
80 Development of a Centralized Academic Hotspot ...
0 Distributed Penetration of Wireless Passwords ...
1 EVALUATION OF STANDARD ROUTING PROTOCOL FOR LO...
2 MAXIMIZE WEBSITE VISIBILITY AND BRANDING BY IM...
3 Utilizing Support Vector Machines to Recognize...
4 ANALYSIS OF BRUTE FORCE ATTACKS ON IP ADDRESS ...
Name: title, dtype: object
SCORE BM25 RANKING (Authors) 10 termirip
39 Kunang YN,Ibadi T
47 Kunang YN,Ibadi TI,...
18 Eka PA,Lin YS,Yesi NK
41 Kunang YN,Yadi IZ
80 Yesi NK,Iلمان ZY,Mm M
0 ...,Kunang YN
1 Adhah PJ,Novaria KY
2 Agus S,Dedi S,Novaria KY
3 Agustian D,Basuki KT,Surya NE,...
4 Alam S,Kunang YN
Name: Authors, dtype: object
```

Gambar 13. Hasil *Queries* dengan Nama Authors

Setelah didapat hasil *queries* pertama, selanjutnya akan digunakan *queries* kedua. Untuk tahapan pencarian *query* kedua sama seperti yang pertama. *Query* kedua yaitu “internet” dengan hasil sebagai berikut.

```
SCORE BM25 RANKING id 10 termirip
[8, 35, 3, 0, 2, 3, 4, 5, 6, 7]

SCORE BM25 RANKING Title 10 termirip
8 IMPLEMENTATION OF COMPUTER DEVICE SALES INFORM...
35 Improving Classification Attacks in IOT Intrus...
1 EVALUATION OF STANDARD ROUTING PROTOCOL FOR LO...
0 Distributed Penetration of Wireless Passwords ...
2 MAXIMIZE WEBSITE VISIBILITY AND BRANDING BY IM...
3 Utilizing Support Vector Machines to Recognize...
4 ANALYSIS OF BRUTE FORCE ATTACKS ON IP ADDRESS ...
5 Application of the Regression Model for Weathe...
6 Implementation of Data Mining Techniques to Pr...
7 Application of Naive Bayes in Malware Detectio...
Name: title, dtype: object
SCORE BM25 RANKING (Authors) 10 termirip
8 Aresvanja E,Kunang YN,St AH,Kom S,Seprina I,...
35 Kunang YN,Nurmaini S,Stiawan D,...
1 Adhah PJ,Novaria KY
0 ...,Kunang YN
2 Agus S,Dedi S,Novaria KY
3 Agustian D,Basuki KT,Surya NE,...
4 Alam S,Kunang YN
5 Amin ME,Kunang YN,...
6 Andri A,Kunang YN,Murniati S
7 Anggraini I,Kunang YN,Firdaus F
Name: Authors, dtype: object
```

Gambar 14. Hasil *Queries* "internet" dengan Nama Authors

Untuk *query* dari *wireless* yang memiliki nilai terbesar adalah dokumen nomor 39 dengan skor BM25 sebesar 3,865 atas nama Yesi Novaria Kunang dan Taqrim Ibadi. untuk *query* kedua didapat nomor dokumen 8 dengan skor BM25 sebesar 5,239 atas nama Fatoni dan Suryayusra.

4. KESIMPULAN DAN SARAN

Dari penelitian yang sudah dilakukan dan dengan dua kali percobaan memasukkan *query* yang berbeda didapat hasil yang berbeda. Percobaan pertama Skor BM25 yang didapat sebesar 3,865 dengan nama *author* Yesi Novari Kunang dan Taqrim Ibadi. Dan untuk percobaan kedua didapat hasil skor BM25 sebesar 5,239 dengan nama *author* Fathoni dan Suryayusra. Dengan hasil pembahasan yang telah diuraikan didapat hasil yang sesuai dengan tujuan penelitian peneliti yaitu mendapatkan hasil rekomendasi dosen pembimbing. Yang menjadi fokus penelitian ini adalah hasil akhir berupa nama dosen yang sesuai dengan topik. Penelitian ini sudah mendapatkan hasil rekomendasi nama dosen dengan ketentuan nilai skor BM25 yang paling tinggi dan berada

diurutan atas maka dosen tersebut berkemungkinan besar menjadi pembimbing untuk mahasiswa yang mengajukan judul/topik.

Dari hasil penelitian yang didapat, saran untuk peneliti lain agar nantinya dapat meneliti kembali dengan metode yang berbeda dan data yang lebih banyak agar nanti hasil yang didapatkan akan jauh lebih baik lagi. Bagi peneliti yang melanjutkan penelitian ini dengan menggunakan metode yang sama diharapkan nantinya agar dapat melakukan uji model supaya nantinya tingkat akurasi dari program yang dibuat mampu dipakai dan juga diharapkan dapat membuat sistem menjadi sebuah aplikasi website ataupun mobile.

Referensi

- ‘Andayani and Ryansyah - 2017 - Implementasi Algoritma TF-IDF Pada Pengukuran Kesa.pdf’ (no date).
- Arisetiawan, A.A.B. (2019) ‘Sistem Rekomendasi Dosen Pembimbing Berdasarkan Dokumen Judul Skripsi di Bidang Komputasi Cerdas Menggunakan Metode BM25’, p. 5.
- Efendi, Z. and Mustakim (2017) *Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi*.
- Mustaqim, T., Umam, K. and Muslim, M.A. (2020) ‘Twitter text mining for sentiment analysis on government’s response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm’, *Journal of Physics: Conference Series*, 1567(3), p. 032024. Available at: <https://doi.org/10.1088/1742-6596/1567/3/032024>.
- Najjichah, H., Syukur, A. and Subagyo, H. (2019) ‘PENGARUH TEXT PREPROCESSING DAN KOMBINASINYA PADA PERINGKAS DOKUMEN OTOMATIS TEKS BERBAHASA INDONESIA’, 15.
- Pardede, J., Husada, M.G. and Riansyah, R. (2021) ‘IMPLEMENTASI DAN PERBANDINGAN METODE OKAPI BM25 DAN PLSA PADA APLIKASI INFORMATION RETRIEVAL’, p. 10.
- Pranata, A.I.B. (2019) ‘Klasifikasi Dokumen pada Laporan Kepolisian dengan Menggunakan Metode BM25 dan Improved K-Nearest Neighbor (IKNN)’, p. 5.
- Salam, A., Wicaksana, V.P. and Hastuti, K. (2015) ‘SISTEM REKOMENDASI PENENTUAN DOSEN PEMBIMBING TUGAS AKHIR DENGAN MENGGUNAKAN ALGORITMA RABIN-KARP’, 14(3), p. 9.
- Vijayarani, S. and Janani, R. (2016) ‘Text Mining: open Source Tokenization Tools – An Analysis’, *Advanced Computational Intelligence: An International Journal (ACII)*, 3(1), pp. 37–47. Available at: <https://doi.org/10.5121/acii.2016.3104>.
- Yulianti, W., Sutomo, B. and Perdana, A. (2020) ‘SISTEM REKOMENDASI PEMILIHAN DOSEN PEMBIMBING SKRIPSI MENGGUNAKAN METODE WEIGHTED PRODUCT (STUDI KASUS STIMIK DHARMA WACANA METRO)’, p. 6.