

Menentukan Dosen Pembimbing Secara Otomatis Dengan Algoritma Text Mining Dan TF-RF

Ryan Andika

Ilmu Komputer dan Teknologi Informasi, Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: ryanandika2407@gmail.com

Abstrak—Sistem penentuan dosen pembimbing secara otomatis adalah, sistem yang digunakan untuk menentukan dosen pembimbing yang tepat sesuai dengan judul skripsi dan metode yang digunakan mahasiswa untuk menyelesaikan tugas akhir atau skripsi. Tujuan penelitian ini adalah untuk mengembangkan Sistem Informasi penentuan dosen pembimbing secara otomatis sehingga dapat mengumpulkan, menganalisis, dan menyajikan informasi yang diperlukan dalam penyusunan data-data dosen pembimbing yang tepat, efisien waktu, dan akurat. Oleh karena itu demi efisiensi dan efektifitas kerja maka pengambilan keputusan yang tepat sangat diperlukan berdasarkan kriteria-kriteria yang ada dan bobot masing-masing kriteria. Menentukan Algoritma Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Proses dalam text mining meliputi proses tokenisasi, stemming dan filtering. Metode pengumpulan data dengan metode kepustakaan, tahapan pengembangan aplikasi meliputi perancangan proses, perancangan tabel, implementasi dan pengujian sistem. Adapun pengerjaan dalam tugas akhir ini menggunakan studi kasus algoritma Text Mining dan algoritma TF-RF dalam menentukan dosen pembimbing tugas akhir atau skripsi untuk mahasiswa. Penelitian ini dilaksanakan di Universitas Budi Darma, sumber data berasal dari portal mahasiswa dan portal dosen. Pengembangan sistem informasi dilakukan dengan menempuh prosedur pengumpulan dan analisis data pendukung melalui observasi, wawancara dan pencatatan secara langsung. Dari hasil perhitungan bobot dengan algoritma TF-IDF, menunjukkan bahwa pada judul skripsi ke-3 adalah nilai nilai bobot yang paling besar yaitu 5.919, berdasarkan hal tersebut maka judul skripsi “Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1” kategorinya adalah Pengolahan citra dan kriptografi.

Kata Kunci: Dosen Pembimbing; Text Mining; TF-RF

Abstract—The system for determining the supervisor automatically is the system used to determine the right supervisor in accordance with the thesis title and the method used by students to complete the final assignment or thesis. The purpose of this research is to develop an Information System for determining supervisors automatically so that it can collect, analyze, and present the information needed in the preparation of appropriate, time-efficient, and accurate data for supervisors. Therefore, for the sake of work efficiency and effectiveness, it is necessary to make the right decision based on the existing criteria and the weight of each criterion. Determining the Algorithm Text mining is the process of mining data in the form of text where the data source is usually obtained from documents and the goal is to find words that can represent the contents of the document so that an analysis of the connectivity between documents can be carried out. Processes in text mining include tokenization, stemming and filtering processes. Methods of data collection using the library method, the stages of application development include process design, table design, system implementation and testing. The work in this final project uses a case study of the Text Mining algorithm and the TF-RF algorithm in determining the supervisor of the final assignment or thesis for students. This research was conducted at Budi Darma University, the source of the data came from the student portal and the lecturer portal. Information system development is carried out by following the procedures for collecting and analyzing supporting data through direct observation, interviews and recording. From the results of weight calculations using the TF-IDF algorithm, it shows that the title of the 3rd thesis has the largest weight value, namely 5,919, based on this, the title of the thesis "Designing a Duplicate Image Scanner Application Using the SHA1 Method" in the category is Image processing and cryptography.

Keywords: Supervisor; Text Mining; TF-RF

1. PENDAHULUAN

Text Mining yang memiliki definisi menambang data yang berupa teks dimana sumber data biasanya di dapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antardokumen. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*textclustering*). Dalam data *science*, kita sering mendengar istilah *machine learning* dan data *visualization*. Tapi sebenarnya, data *science* mencakup bidang yang jauh lebih luas banyak istilah-istilah Teknologi yang termasuk dalam kategori “*data science*”. Salah satunya yang cenderung lebih jarang kita *Text Mining*, atau dikenal juga dengan istilah *textanalytics*.

Pada dasarnya, *Text Mining* merupakan Teknologi *Artificial Intelligence* (AI) yang memungkinkan penggunaannya untuk mengubah konten inti dari sebuah dokumen teks menjadi sebuah data kuantitatif secara cepat. Data kuantitatif tersebut nantinya akan dapat digunakan atau ditindaklanjuti sesuai keinginan penggunaannya. dalam sebuah kampus menentukan dosen pembimbing. Para admin kampus harus menemukan informasi yang tepat dari sejumlah teks yang sangat besar. Proses manual ini akan menghabiskan waktu yang sangat lama. Menemukan informasi tersebut sangat penting dalam membuat keputusan yang tepat, tapi sekitar 80% informasi tersebut tersembunyi dalam kumpulan teks yang tidak beraturan. Dengan menggunakan metode *keywordsearch*

konvensional, tenaga manusia hanya dapat menarik sejumlah dokumen yang mengandung keyword tersebut, dan masih harus membacanya dengan teliti.

Dengan bantuan Teknologi *Text Mining*, proses tersebut dapat menjadi jauh lebih efisien dibandingkan proses manual, dan juga menarik informasi atau wawasan yang mungkin tidak bisa didapatkan secara manual. Informasi tersebut kemudian dapat diolah menjadi data terstruktur untuk dianalisa dan divisualisasikan. Intinya, penggunaan *text mining* dapat membantu kita untuk meningkatkan efisiensi analisis dokumen teks dengan memberikan kemampuan otomatisasi pada proses tersebut.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Windu Gata, Purnomo yang dipublikasi pada jurnal Format Vol. 6 No 1 Tahun 2017 Sebuah penelitian yang mengevaluasi keefektifan algoritma K-Nearest Neighbour (KNN) dalam mengklasifikasikan konten berita SMS dengan menentukan tingkat akurasi text mining. Hasil penelitian yang dilakukan mendapatkan hasil pada akurasi seleksi prediksi sejumlah 772 benar dan tidak sesuai jumlah 32, sehingga presisinya adalah 96.02%. Sedangkan prediksi tidak mempunyai hasil sejumlah 0 kesalahan dan 14 benar dalam prediksi. Sehingga hasil akurasi yang didapatkan sejumlah 96.15% [1].

Penelitian yang sama dilakukan oleh Agatha Deolika, Kusri, Emha Taufiq Luthf iyang dipublikasi pada jurnal Teknologi Informasi Vol. 3 No 2 Tahun 2019, dimana melakukan klasifikasi pembobotan kata dengan text mining. Berdasarkan analisa pada hasil klasifikasi yang telah dilakukan, dapat disimpulkan bahwa pembobotan TF-RF dengan klasifikasi *Naïve bayes* lebih baik dari pembobotan TF-IDF dan WIDF dengan nilai Accuracy 98,67%, Precision 93,81%, dan Recall 96,67%. Klasifikasi *naïve bayes* dapat digunakan untuk mengelompokkan atau klasifikasi text mining dengan baik [2].

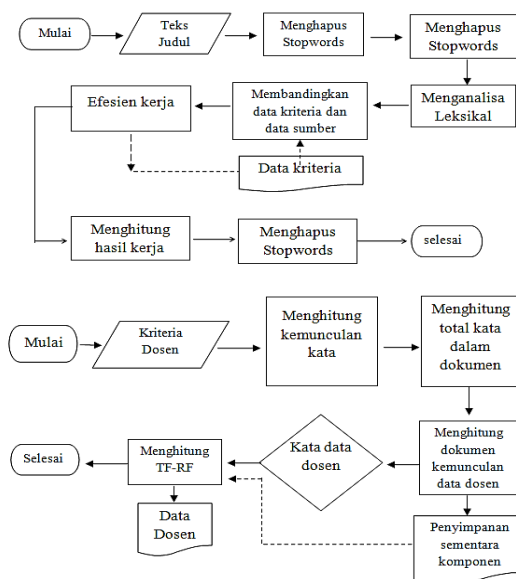
Dengan bantuan Teknologi *Text Mining*, proses tersebut dapat menjadi jauh lebih efisien dibandingkan proses manual, dan juga menarik informasi atau wawasan yang mungkin tidak bisa didapatkan secara manual. Informasi tersebut kemudian dapat diolah menjadi data terstruktur untuk dianalisa dan divisualisasikan. Intinya, penggunaan *text mining* dapat membantu kita untuk meningkatkan efisiensi analisis dokumen teks dengan memberikan kemampuan otomatisasi pada proses tersebut.

Metode TF-RF yang berasal dari TF-IDF, TF-RF adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing- masing *term* di asumsikan mempunyai proporsi kepentingan sesuai jumlah terjadinya (munculnya) dalam teks (dokumen). *Term frequency* dapat memperbaiki nilai *recall* pada informasi *retrieval*, tetapi tidak selalu memperbaiki nilai *precision*. Hal ini disebabkan *term* yang *frequent* cenderung muncul di banyak teks, sehingga *term-term* tersebut memiliki kekuatan pembeda (keunikan) yang kecil. Sedangkan *term frequency* (TF) adalah pengukuran yang paling sederhana dalam metode pembobotan. Pada metode ini, masing-masing *term* di asumsikan mempunyai proporsi kepentingan sesuai jumlah terjadinya (munculnya) dalam teks (dokumen). *Term frequency* dapat memperbaiki nilai *recall* pada information retrieval, tetapi tidak selalu memperbaiki nilai *precision*.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Untuk tahapan-tahapan secara lengkap penerapan algoritma text mining dalam menentukan pemilihan dosen pembimbing pada kampus dapat dilihat pada gambar 1 berikut.



Gambar 1. Kerangka Kerja Penelitian

2.2 String Matching

String matching adalah suatu algoritma yang digunakan untuk memecahkan masalah pencocokan suatu teks terhadap suatu teks lain. Ada beberapa algoritma yang dapat digunakan dalam pencarian string matching contohnya ialah algoritma Boyer Moore dan Brute Force. Algoritma Boyer-Moore adalah salah satu algoritma pencarian string, dipublikasikan oleh Robert S. Boyer, dan J. Strother Moore pada tahun 1977. Algoritma ini dianggap sebagai algoritma yang paling efisien pada aplikasi umum. Tidak seperti algoritma pencarian string yang ditemukan sebelumnya, algoritma Boyer-Moore mulai mencocokkan karakter dari sebelah kanan pattern (pola yang dicari). Algoritma Brute Force adalah algoritma untuk mencocokkan pattern dengan semua teks antara 0 dan n-m untuk menemukan keberadaan pattern dalam teks. Maka perlu adanya analisa yang membandingkan algoritma Boyer Moore dan Brute Force pada pencarian Kamus Besar Bahasa Indonesia berbasis android. Hasil dari penelitian ini menunjukkan bahwa algoritma Brute Force lebih cepat dibandingkan algoritma Boyer Moore pada pencarian Kamus Besar Bahasa Indonesia, Hasil total running time algoritma Brute Force adalah 168,3 ms pada kata, 6994,16ms pada deskripsi dan algoritma Boyer Moore adalah 304,7 ms pada kata, 8654,77 ms pada deskripsi. Untuk pengujian keyword related kedua algoritma dapat menampilkan daftar keyword related yang sama. Pada algoritma Boyer Moore maupun Brute Force, semakin sedikit karakter pattern yang dicari maka pencariannya semakin lama karena yang dicari semakin banyak dan kata yang ditemukan juga semakin banyak, dan semakin banyak karakter pattern yang dicari maka pencariannya semakin cepat karena yang dicari semakin sedikit dan kata yang ditemukan juga semakin sedikit [3][4][5].

2.3 Text Mining

Text mining adalah proses untuk memperoleh informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik. Pada proses teks mining terdapat pembobotan kata yang bertujuan untuk memberikan nilai/bobot pada term yang terdapat pada suatu dokumen. Bobot yang diberikan pada term tergantung kepada metode yang digunakan. Dalam pembobotan kata banyak sekali terdapat algoritma-algoritma contohnya seperti TF, Idf, RF, TF-IDF, TF.RF, WIDF [6][7][8].

Text Mining merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi, dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. Dan juga *text mining* dapat dikatakan sebagai penerapan konsep dari teknik data mining untuk mencari pola dalam teks, bertujuan untuk mencari informasi yang bermanfaat dengan tujuan tertentu[9][10][11].

2.4 Term Frequency-Relevance Frequency (TF-RF)

TF-RF adalah metode gabungan antara TF dan RF dengan tujuan untuk mendapatkan performansi yang lebih baik dari metode pembobotan kata lainnya. Metode ini mempertimbangkan relevansi dokumen dilihat dari frekuensi kemunculan *term* di kategori yang berkaitan. Persamaan metode TF-RF sebagai berikut[12][13][14]:

$$TF * RF(t, c) = TF(d, t) * \log_2 \left(2 + \frac{a}{\max(1, c)} \right) \quad (1)$$

Dimana :

$TF * RF(d, t)$: Pembobotan TF-RF.

$TF(d, t)$: Frekuensi munculnya *term* pada dokumen *d*.

c : Kelas kategori.

t : *term*.

2.5 Dosen Pembimbing

Dosen Pembimbing adalah dosen yang ditetapkan untuk melaksanakan pembimbingan kepada mahasiswa yang berfungsi sebagai sumber informasi, sebagai penampung kesulitan, sebagai petunjuk dan pemberi saran serta sebagai pemantau kegiatan mahasiswa. Usaha-usaha yang dilakukan oleh dosen pembimbing akademik terhadap mahasiswa bimbingan yang dilakukan secara teratur dan terarah sehingga dapat memotivasi kegiatan belajar sehingga mahasiswa akan dapat mencapai prestasi belajar yang maksimal. Pembimbingan akademik secara umum meliputi kegiatan (1) membina dan mengarahkan mahasiswa agar dapat bersikap sebagai ilmuwan dalam rangka mengembangkan kebebasan akademik sesuai dengan bidang ilmu yang ditempuhnya, (2) menentukan jumlah beban studi yang akan diambil mahasiswa untuk semester yang sedang berjalan, (3) memantau dan membimbing mahasiswa demi kelancaran studinya serta membantu memecahkan masalah-masalah yang dihadapi mahasiswa yang bersangkutan baik yang bersifat akademik maupun nonakademik yang diperkirakan dapat mengganggu pencapaian keberhasilan studi, dan (4) menentukan/memutuskan boleh atau tidaknya mahasiswa yang bersangkutan menempuh ujian akhir semester [15].

3. HASIL DAN PEMBAHASAN

Pada penelitian ini data-data dosen akan di proses untuk menghasilkan pemilihan dosen pembimbing secara otomatis dengan menggunakan Algoritma *Text Mining* dan TF-RF. Untuk menyelesaikan masalah diatas penulis menerapkan algoritma *Text Mining* yang merupakan suatu teknik data mining untuk mencari pola dalam penentuan kriteria dosen pada kampus. Proses penganalisaan teks guna menemukan informasi yang bermanfaat untuk tujuan tertentu. *Algoritma text mining* memiliki tahapan *case folding*, *tokenizing*, *filtering*, *stemming*, *spelling normalization*, dan *analyzing* hasil dari tahapan-tahapan adalah kata-kata *root* atau kata-kata akar dari suatu kalimat.

3.1 Penerapan Algoritma *Text Mining* dan TF-RF

Tahapan awal yang dilakukan pada penelitian ini adalah menentukan data awal sebagai pegetahuan awal, dalam penelitian tersebut peneliti telah mengambil data dosen - dosen beserta kategorinya sebanyak 5 kriteria yang dapat dilihat pada tabel 1.

Tabel 1. Data Awal Judul Skripsi

No.	Judul Skripsi	Kategori
1.	Sistem Pendukung Keputusan Pemilihan Jenis Lampu untuk Pencahayaan Ruang Menggunakan Metode AHP	Sistem Keputusan Pendukung
2.	Pengembangan Infrastruktur Jaringan Menggunakan Metode Sekmentasi Di Universitas Mulia	Jaringan
3.	Sistem Pendukung Keputusan Pengadaan Raw Material Pembuatan Mie Instan Menggunakan Metode Analithical Hierarchy Process Studi Kasus PT. Indofood CBP Sukses Makmur	Sistem Keputusan Pendukung
4.	Implementasi Metode Fuzzy Tsukamoto Dalam Menentukan Keluarga Penerima Bedah Rumah Di Desa Onanunggu LV Kec.Sipahutar	String Matching
5.	Implementasi Kriptografi Dan Steganografi Menggunakan Algoritma RSA Dan Metode LSB	Kriptografi
6.	Penerapan Algoritma Knuth Morris Pratt Pada Aplikasi Mercury Untuk Pencarian Data Merchant (Studi Kasus : PT. Visionet Data International)	String Matching
7.	Impelementasi Algoritma Text Mining TF-IDF Untuk Fitur Autoresponder	String Matcing
8.	Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1	Pengolahan Kriptografi Citra,
9.	Penerapan Algoritma RSA Dan DES Pada Pengamanan File Teks	Kriptografi
10.	Permainan <i>World Search Puzzle</i> Pada Android Menggunakan Algoritma <i>Boyer Moore</i>	String Matching

Setelah data awal tentukan maka dilakukan pengolahan teks sesuai dengan algoritma text mining, adapun tahapan dari algoritma text mining adalah berikut :

1. *Case Folding*

Case Folding adalah tahapan selanjutnya setelah tahapan *cleaning* dimana sistem mengubah semua huruf dalam teks menjadi huruf kecil. Karena tidak semua kata dalam teks menggunakan huruf kapital oleh karena itu peran *case folding* dibutuhkan untuk mengkonversi keseluruhan teks menjadi sebuah bentuk standar sehingga memudahkan untuk melanjutkan ke tahapan *tokenizing* selanjutnya. Hasil dari tahapan *case folding* dapat dilihat pada tabel 2 berikut:

Tabel 2. Tahap *Case Folding*

No.	Judul Skripsi	Kategori
1.	Sistem pendukung keputusan pemilihan jenis Lampu untuk pencahayaan ruangan menggunakan metode ahp	Sistem Keputusan Pendukung
2.	Pengembangan infrastruktur jaringan menggunakan metode sekmentasi di universitas mulia	Jaringan
3.	Sistem pendukung keputusan pengadaan raw material pembuatan mie instan menggunakan metode analithical hierarchy process studi kasus pt. Indofood cbp sukses makmur	Sistem Keputusan Pendukung
4.	Implementasi metode fuzzy tsukamoto dalam menentukan keluarga penerima bedah rumah di desa Onanunggu lv kec.sipahutar	String Matching

No.	Judul Skripsi	Kategori
5.	Implementasi kriptografi dan steganografi menggunakan algoritma rsa dan metode lsb	Kriptografi
6.	Penerapan algoritma knuth morris pratt pada aplikasi mercury untuk pencarian data merchant (studi Kasus : pt. Visionet data international)	String Matching
7.	Impelementasi algoritma text mining tf-idf untuk fitur autoresponder	String Matcing
8.	Perancangan aplikasi duplicate image scanner menerapkan metode sha1	Pengolahan Citra, Kriptografi
9.	Penerapan algoritma rsa dan des pada pengamanan file teks	Kriptografi
10.	Permainan world search puzzle Pada Android Menggunakan algoritma boyer moore	String Matching

2. *Tokenizing*

Tokenizing adalah tahap pemotongan string input berdasarkan tiap kalimat yang menyusunnya. Tokenisasi secara garis besar memecah kalimat dalam suatu teks ke dalam bentuk sekumpulan satuan kalimat untuk dihitung hasilnya nanti. Proses tokenisasi akan semakin sulit dikarenakan harus memperhatikan struktur bahasa. Tahapan *Tokenizing* dapat dilihat seperti pada tabel 3 berikut :

Tabel 3. Tahapan *Tokenizing*

No.	Judul Skripsi	Kategori
1.	Sistem pendukung keputusan pemilihan jenis Lampu pencahayaan ruangan menggunakan metode ahp	Sistem Pendukung Keputusan
2.	Pengembangan infrastruktur jaringan menggunakan metode sekmentasi	Jaringan
3.	Sistem pendukung keputusan pengadaan raw material pembuatan mie instan menggunakan metode analithical hierarchy process	Sistem Pendukung Keputusan
4.	Implementasi metode fuzzy tsukamoto dalam menentukan keluarga penerima bedah rumah	String Matching
5.	Implementasi kriptografi dan steganografi menggunakan algoritma rsa dan metode lsb	Kriptografi
6.	Penerapan algoritma knuth morris pratt pada aplikasi mercury pencarian data merchant	String Matching
7.	Impelementasi algoritma text mining tf-idf fitur autoresponder	String Matcing
8.	Perancangan aplikasi duplicate image scanner menerapkan metode sha1	Pengolahan Citra, Kriptografi
9.	Penerapan algoritma rsa dan des pengamanan file teks	Kriptografi
10.	Permainan world search puzzle pada android Menggunakan algoritma boyer moore	String Matching

3. *Filtering*

Pada tahapan ini akan dilakukan penghapusan kata yang tidak penting, pada tahapan tersebut untuk memudahkan penghapusan kata tidak penting penulis memanfaatkan library stopwords dari tala. Hasil dari tahapan tersebut dapat dilihat pada tabel 4.

Tabel 4. Tahapan *Filtering*

No.	Judul Skripsi	Kategori
1.	Sistem pendukung keputusan pemilihan jenis Lampu pencahayaan ruangan menggunakan metode ahp	Sistem Pendukung Keputusan
2.	Pengembangan infrastruktur jaringan menggunakan metode sekmentasi	Jaringan
3.	Sistem pendukung keputusan pengadaan raw material pembuatan mie instan menggunakan metode analithical hierarchy process	Sistem Pendukung Keputusan
4.	Implementasi metode fuzzy tsukamoto dalam menentukan keluarga penerima bedah rumah	String Matching
5.	Implementasi kriptografi dan steganografi menggunakan algoritma rsa dan metode lsb	Kriptografi
6.	Penerapan algoritma knuth morris pratt pada aplikasi mercury pencarian data merchant	String Matching

No.	Judul Skripsi	Kategori
7.	Impelementasi algoritma text mining tf idf fitur autoresponder	String Matcing
8.	Perancangan aplikasi duplicate image scanner menerapkan metode sha1	Pengolahan Citra, Kriptografi
9.	Penerapan algoritma rsa dan des pengamanan file teks	Kriptografi
10.	Permainan world search puzzle pada android Menggunakan algoritma boyer moore	String Matching

4. *Stemming*

Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Hasil dari tahapan tersebut dapat dilihat pada tabel 5.

Tabel 5. Tahapan *Stemming*

No.	Judul Skripsi	Kategori
1.	Sistem pendukung keputusan pemilihan jenis Lampu pencahayaan ruangan menggunakan metode ahp	Sistem Pendukung Keputusan
2.	Pengembangan infrastruktur jaringan menggunakan metode sekmentasi	Jaringan
3.	Sistem pendukung keputusan pengadaan raw material pembuatan mie instan menggunakan metode analithical hierarchy process	Sistem Pendukung Keputusan
4.	Implementasi metode fuzzy tsukamoto dalam menentukan keluarga penerima bedah rumah	String Matching
5.	Implementasi kriptografi dan steganografi menggunakan algoritma rsa dan metode lsb	Kriptografi
6.	Penerapan algoritma knuth morris pratt pada aplikasi mercury pencarian data merchant	String Matching
7.	Impelementasi algoritma text mining tf idf fitur autoresponder	String Matcing
8.	Perancangan aplikasi duplicate image scanner menerapkan metode sha1	Pengolahan Citra, Kriptografi
9.	Penerapan algoritma rsa dan des pengamanan file teks	Kriptografi
10.	Permainan world search puzzle pada android Menggunakan algoritma boyer moore	String Matching

5. *Analyzing*

Analyzing Merupakan tahap penentuan seberapa jauh keterhubungan antar kata-kata dokumen yang ada. Untuk memudahkan analisa keterbuhubungan antar kata-kata pada setiap nama dosen yang telah diketahui kategorinya nama dosen yang belum diketahui kategorinya, maka penulis menerapkan algoritma *Term Frequency – Relevance Frequency* (TF-RF). Algoritma tersebut akan membantu penulis dalam menganalisa hubungan antara masing-masing kata dari setiap nama dosen.

Pada penelitian tersebut penulis telah menentukan nama dosen yang akan ditentukan kategorinya yaitu “Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1”. Untuk memudahkan proses perbandingan antara judul skripsi yang sudah diketahui kategorinya dengan judul skripsi yang belum diketahui kategorinya maka penulis melakukan tahapan algoritma text mining terhadap judul skripsi yang belum diketahui kategorinya tersebut. Hasil dari penerapan algoritma text mining terhadap judul skripsi yang belum diketahui kategorinya tersebut adalah “Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1”.Setelah kedua judul diterapkan algoritma text mining maka selanjutnya melakukan perbandingan atas kedua judul tersebut dengan menerapkan algoritma *Term Frequency* (TF-RF). Hasil dari algoritma tersebut dapat dilihat pada tabel 6.

Tabel 6. Hasil Perhitungan Term Frequency

Q	Tf										df	D/df	IDF	IDF+1
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10				
Aman	1	0	0	1	0	0	0	1	1	0	4	2,5	0,397	1,397
Teks	0	1	0	0	0	0	0	0	1	0	2	5	0,698	1,698
Enkripsi	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Algoritma rc4+	0	0	1	1	1	0	1	0	1	0	5	2	0,301	1,301
steganografi	0	0	1	0	1	0	0	1	0	1	4	2,5	0,397	1,397
Dcs	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Q	Tf										df	$\frac{D}{df}$	IDF	IDF+1
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10				
Citra	2	0	1	0	0	0	0	0	0	0	2	5	0,698	1,698
Digital	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabel 7. Hasil Pembobotan TF-RF

W=tf*(IDF+1)										
d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	
1,397	0	0	1397	0	0	0	1397	1,397	0	0
0	1,698	0	0	0	0	0	0	1,698	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	1,301	1,301	1,301	0	1,301	0	1,301	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	1,397	0	1,397	0	0	1,397	0	1,397	0
0	0	0	0	0	0	0	0	0	0	0
3,396	0	1,698	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
sum(d1)	sum(d2)	sum(d3)	sum(d4)	sum(d5)	sum(d6)	sum(d7)	sum(d8)	sum(d9)	sum(d10)	
4,793	1,698	5,919	2,698	2,698	0	1,301	2,794	4,396	1,397	

Dari hasil perhitungan bobot dengan algoritma TF-IDF, menunjukkan bahwa pada judul skripsi ke-3 adalah nilai nilai bobot yang paling besar yaitu 5.919, berdasarkan hal tersebut maka judul skripsi “Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1” kategorinya adalah Pengolahan citra dan kriptografi.

Dari hasil perhitungan diatas, jika dihubungkan dengan tabel daftar dosen dibawah ini dapat terlihat jelas bahwa dosen yang sesuai untuk judul “Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1” cocok dengan Dosen Pembimbing Pristiwanto, M.Kom dan Dosen Pembimbing Saidi Ramadhan Siregar, M.Kom.

Tabel 8. Daftar Nama Dosen Pembimbing

No.	Nama	Kategori
1.	Guidio L Ginting, M.Kom	String Matching
2.	Efori Buulolo, M.Kom	Data Mining
3.	Pristiwanto, M.Kom	Pengolahan Citra
4.	Saidi Ramadhan Siregar, M.Kom	Kriptografi
5.	Riswan Limbong, M.Kom	Steganografi
6.	Pilipus Tarigan, M.Kom	Pemodelan dan Simulasi
7.	Nelly Astuti Hasibuan, M.Kom	Sistem Pakar
8.	Hery Sunandar, M.Kom	Jaringan
9.	Mesran, M.Kom	SPK
10.	Garuda Ginting, M.Kom	Natural Language Processing

3.2 Hasil Pengujian

1. Tampilan Menu Utama

Menu utama adalah menu yang menyediakan pilihan untuk menginput data ke tujuan, input data untuk penentuan dosen pembimbing, dan menu keluar

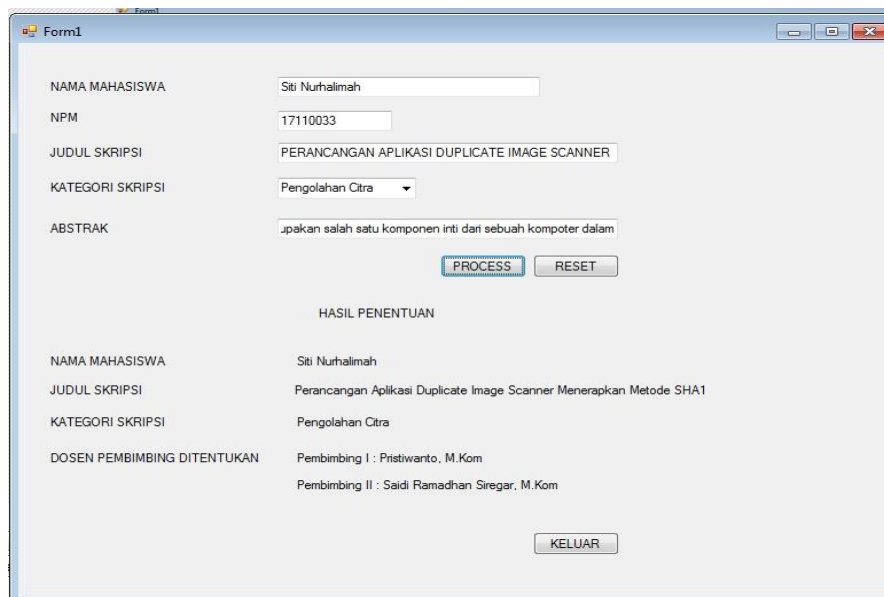
The screenshot shows a web form titled 'Form1'. It contains several input fields: 'NAMA MAHASISWA', 'NPM', 'JUDUL SKRIPSI', 'KATEGORI SKRIPSI' (with a dropdown arrow), and 'ABSTRAK'. Below these fields are two buttons: 'PROCESS' and 'RESET'. Underneath, there is a section titled 'HASIL PENENTUAN' which contains four more input fields: 'NAMA MAHASISWA', 'JUDUL SKRIPSI', 'KATEGORI SKRIPSI', and 'DOSEN PEMBIMBING DITENTUKAN'. At the bottom right of this section is a button labeled 'KELUAR'.

Gambar 2. Tampilan Form Menu Utama

Keterangan :

- a. Menu Proses Identitas dan Kategori Dosen memiliki 2 *sub menu* yaitu *sub menu form* dan sub form keluar..
 - b. Menu Keluar berfungsi untuk keluar dari *system* Perhitungan Dosen Pembimbing.
2. Tampilan Form Hasil

Tampilan *form* hasil merupakan tampilan hasil dari menentukan dosen pembimbing secara otomatis yang muncul saat pengguna telah selesai melakukan proses *input* data. Berikut ini adalah gambar dari layar *Form* hasil, seperti pada gambar berikut:



The screenshot shows a web form with the following fields and content:

NAMA MAHASISWA	Siti Nurhalimah
NPM	17110033
JUDUL SKRIPSI	PERANCANGAN APLIKASI DUPLICATE IMAGE SCANNER
KATEGORI SKRIPSI	Pengolahan Citra
ABSTRAK	.upakan salah satu komponen inti dari sebuah kompoter dalam

Buttons: PROCESS, RESET

HASIL PENENTUAN

NAMA MAHASISWA	Siti Nurhalimah
JUDUL SKRIPSI	Perancangan Aplikasi Duplicate Image Scanner Menerapkan Metode SHA1
KATEGORI SKRIPSI	Pengolahan Citra
DOSEN PEMBIMBING DITENTUKAN	Pembimbing I : Pristiwanto, M.Kom Pembimbing II : Saidi Ramadhan Siregar, M.Kom

Button: KELUAR

Gambar 3. Tampilan Form Hasil

4. KESIMPULAN

Setelah dilakukan pengujian dengan menggunakan metode Text Mining dan TF-RF dalam penentuan Dosen Pembimbing Secara Otomatis, penulis dapat mengambil kesimpulan yang diharapkan dapat bermanfaat bagi banyak bidang terutama dalam bidang pendidikan. Adapun kesimpulannya yaitu metode Text Mining dan TF-RF dalam menentukan dosen pembimbing dengan bahasa pemograman dalam hal ini penulis menggunakan Visual Basic 2008 sebagai alat bantu bisa dapat dijalankan sesuai dengan yang diinginkan, dan mendapatkan program yang dapat berjalan. Algoritma Text Mining dan TF-RF dapat diterapkan dalam program from Identitas dan Kategori Dosen untuk menentukan Dosen Pembimbing pada mata kuliah yg telah ditentukan.

REFERENCES

- [1] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, pp. 1–13, 2017.
- [2] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [3] A. Januardi, "Analisa Perbandingan Algoritma Brute Force Dan Boyer Moore Dalam Pencarian Word Suggestion Eksponensial," *Pelita Inform. Budi Darma*, vol. IV, pp. 18–24, 2013.
- [4] M. N. Fauzy and K. Kusriani, "Chatbot menggunakan metode fuzzy string matching sebagai virtual assistant pada pusat layanan informasi akademik," *J. Inf. J. Penelit. dan Pengabd. Masy.*, vol. 5, no. 1, pp. 61–67, 2019.
- [5] E. Ernawati, A. Johar, and S. Setiawan, "Implementasi Metode String Matching Untuk Pencarian Berita Utama Pada Portal Berita Berbasis Android (Studi Kasus: Harian Rakyat Bengkulu)," *Pseudocode*, vol. 6, no. 1, pp. 77–82, 2019.
- [6] A. Priyanto and M. R. Ma'arif, "Implementasi web scrapping dan text mining untuk akuisisi dan kategorisasi informasi dari internet (studi kasus: Tutorial hidroponik)," *Indones. J. Inf. Syst.*, vol. 1, no. 1, pp. 25–33, 2018.
- [7] K. Fitriani, I. Isbandi, and A. Amaliyah, "Perancangan Sistem Manajemen Dokumen Dengan Menggunakan Metode Text Mining Di Kantor Kelurahan Sekejati," *Telematika*, vol. 3, no. 1, pp. 45–59, 2021.
- [8] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid-19 Menggunakan Metode Naïve Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021.
- [9] E. Yulian, "Text mining dengan k-means clustering pada tema LGBT dalam arsip tweet masyarakat Kota Bandung," *J. Mat. "MANTIK"*, vol. 4, no. 1, pp. 53–58, 2018.
- [10] A. F. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi:(Sebuah Ulasan)," *JUPITER J. Penelit. Ilmu dan Teknol. Komput.*, vol. 13, no. 1, pp. 66–78, 2021.
- [11] A. V. Sudiantoro and E. Zuliarso, "Analisis sentimen twitter menggunakan text mining dengan algoritma Naïve Bayes Classifier," 2018.
- [12] R. Ramadhan, Y. A. Sari, and P. P. Adikara, "Perbandingan Pembobotan Term Frequency-Inverse Document Frequency

- dan Term Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5075–5079, 2021.
- [13] F. I. N. Haq and E. B. Setiawan, “Implementasi Naive Bayes Classifier untuk Prediksi Kepribadian Big Five pada Twitter Menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Term Frequency-Relevance Frequency (TF-RF),” *eProceedings Eng.*, vol. 6, no. 2, 2019.
- [14] M. F. Luthfi and K. M. Lhaksamana, “Implementation of TF-IDF method and support vector machine algorithm for job applicants text classification,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1181–1186, 2020.
- [15] D. Y. KHAIRUN and I. AL HAKIM, “Profil Layanan Dosen Pembimbing Akademik,” *J. EDUKASI J. Bimbing. Konseling*, vol. 4, no. 1, p. 31, 2018, doi: 10.22373/je.v4i1.3521.