

Optimasi Penerimaan Siswa Baru dengan Penerapan Algoritma Text Mining dan TF-IDF

Y Ayu Putri Gabriella S*, Guidio L Ginting, Melda Panjaitan

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

Email: ^{1,*}itsmeayuputri12@gmail.com, ²guidio.leonard626@gmail.com, ³meldapjt.78@gmail.com

Email Penulis Korespondensi: itsmeayuputri12@gmail.com

Abstrak—Sekolah Menengah Pertama Yayasan Perguruan merupakan salah satu sekolah yang banyak diminati oleh beberapa kalangan, dikarenakan jurusan yang bermacam-macam yang ada pada sekolah tersebut. Pada saat ini sekolah tersebut masih terbilang banyak memiliki permasalahan dalam penerimaan siswa baru nya. Banyak data yang masuk tidak sesuai dengan apa yang telah diterima, dengan demikian sekolah perlu menerapkan perancangan aplikasi dalam penerimaan siswa baru agar dokumen dan data yang diterima lebih konkret, tidak hanya itu dengan menerapkan metode tersebut, pengerjaan dan proses pemasukan data dapat memberikan hasil yang maksimal. Metode yang digunakan dalam penelitian ini yaitu *text mining* dan TF-IDF, *text mining* sendiri yaitu metode yang memiliki tahapan untuk menemukan data menjadi lebih efektif dan akurat. Selain itu metode TF-IDF pun memiliki cara pengerjaan yang dapat dikatakan efisien dengan cara menghitung bobot pada setiap kata. Sistem penerimaan siswa baru ini dibangun menggunakan software *dreamweaver*, dan bahasa pemrograman HTML, CSS, dan *JavaScript* serta menggunakan database MySQL sebagai database server. Hasil dari penelitian menggunakan metode ini dapat memberikan kemudahan pihak sekolah dan calon siswa dalam penerimaan siswa baru yang dilakukan secara *online* maupun *offline*, sehingga pendataan, dan penerimaan siswa baru dapat dilakukan dengan sistem otomatis yang tidak perlu menggunakan cara kerja yang bertele-tele.

Kata Kunci: Optimasi; Text Mining; TF-IDF

Abstract—The North Sumatra Model College Foundation Junior High School is one of the schools that is in great demand by several groups, due to the various majors that exist at the school. At this time the school still has many problems in admitting new students. A lot of incoming data does not match what has been received, thus schools need to implement application design in new student admissions so that the documents and data received are more concrete, not only that by applying this method, the processing and data entry process can provide maximum results. The methods used in this research are text mining and TF-IDF, text mining itself is a method that has stages to find data to be more effective and accurate. In addition, the TF-IDF method also has an efficient way of working by calculating the weight of each word. This new admissions system is built using Dreamweaver software, and HTML, CSS, and JavaScript programming languages and uses a MySQL database as a database server. The results of research using this method can make it easier for schools and prospective students to accept new students online and offline, so that data collection and new student admissions can be carried out with an automatic system that does not need to use long-winded ways of working.

Keywords: Optimization; Text Mining; TF-IDF

1. PENDAHULUAN

Teknologi informasi telah menjadi bagian integral dari kehidupan sehari-hari masyarakat, dan perkembangannya terus meningkat seiring dengan kemajuan globalisasi[1]. Teknologi informasi telah membawa perubahan signifikan di berbagai sektor, termasuk transportasi, komunikasi, kesehatan, pendidikan, dan banyak lagi. Salah satu kemajuan yang paling mencolok adalah akses mudah ke informasi melalui internet. Ini memungkinkan pengguna internet untuk dengan cepat mengakses informasi yang mereka butuhkan[2].

Sekolah Yayasan Perguruan adalah salah satu institusi pendidikan yang telah beradaptasi dengan perkembangan teknologi informasi. Mereka menawarkan berbagai program keahlian, termasuk Teknik Otomotif, Teknik Audio Video, Teknik Pemanfaat Tenaga Listrik, Teknik Komputer dan Jaringan, serta Teknik Pemesinan. Visi dan misi sekolah ini didukung oleh kerja keras dari guru, staf, dan orangtua siswa. Setiap tahun, sekolah ini melakukan proses penerimaan siswa baru untuk memilih calon siswa yang akan melanjutkan pendidikan mereka ke jenjang Sekolah Menengah Atas (SMA).

Selama ini, sekolah Yayasan Perguruan telah mengelola data penilaian calon siswa menggunakan Microsoft Excel. Namun, metode ini tidak memadai dalam mengatasi kendala yang muncul, terutama dalam menentukan penilaian terhadap calon siswa. Proses ini memakan waktu yang relatif lama, dan hasil kelulusan calon siswa baru bisa diperoleh dalam satu minggu. Sekolah ini merasa perlu untuk memperbaiki sistem penilaian ini dan mencari metode yang lebih efisien. Mereka ingin suatu metode baru yang dapat mengatasi semua kendala ini, mulai dari proses pemasukan data hingga hasil kelulusan calon siswa, agar informasi yang dihasilkan bermanfaat dan membantu sekolah dalam menentukan kelulusan.

Untuk mengatasi permasalahan ini, sekolah Yayasan Perguruan memutuskan untuk mengadopsi algoritma text mining dan TF-IDF. Algoritma text mining adalah metode yang digunakan untuk menganalisis teks dan mengambil informasi yang berguna dari dokumen-dokumen teks. TF-IDF (Term Frequency-Inverse Document Frequency) adalah salah satu teknik yang digunakan dalam text mining untuk memberikan bobot pada kata-kata dalam dokumen.

Penelitian sebelumnya telah menunjukkan keefektifan algoritma text mining dalam berbagai aplikasi. Sebagai contoh, penelitian oleh Andi Taufik pada tahun 2018 membahas penggunaan algoritma text mining dalam klasifikasi review hotel. Hasil penelitiannya menunjukkan bahwa metode ini efektif dalam mengambil pengetahuan yang akurat dari dokumen teks yang diinginkan[3].

Selain itu, Immanuel Olive Djaja Putra dan rekan-rekannya melakukan penelitian pada tahun 2020 tentang implementasi text mining dalam analisis opini masyarakat terhadap layanan transportasi online. Mereka menggunakan text mining untuk menganalisis data dan mengklasifikasikan informasi, memberikan pandangan yang berharga tentang faktor-faktor yang memengaruhi persepsi masyarakat terhadap layanan tersebut [4].

Dalam konteks sekolah Yayasan Perguruan, algoritma text mining dan TF-IDF dapat digunakan untuk mengelola data penilaian calon siswa dengan lebih efisien. Dengan metode ini, proses penilaian bisa dilakukan lebih cepat dan hasilnya bisa lebih akurat. Data calon siswa dapat dianalisis dengan lebih baik, dan keputusan kelulusan dapat diambil dengan lebih cepat.

Selain itu, metode TF-IDF juga dapat membantu dalam memberikan bobot pada kata-kata yang relevan dalam dokumen-dokumen penilaian. Hal ini dapat membantu dalam mengidentifikasi faktor-faktor yang memengaruhi keputusan kelulusan. Dengan demikian, sekolah dapat mengambil keputusan yang lebih baik dan objektif.

Berdasarkan penelitian yang dilakukan oleh Ria Melita dan rekan-rekannya pada tahun 2018, TF-IDF merupakan suatu teknik yang bermanfaat dalam memberikan bobot pada kata-kata dalam dokumen. Teknik ini dapat digunakan untuk mengukur relevansi kata-kata dengan dokumen tertentu dan membantu dalam pengambilan keputusan [5].

Dengan adopsi algoritma text mining dan TF-IDF, sekolah Yayasan Perguruan dapat mengoptimalkan proses penerimaan siswa baru mereka. Mereka dapat mengelola data dengan lebih efisien, mengambil keputusan kelulusan dengan lebih cepat, dan mendapatkan wawasan yang lebih dalam tentang faktor-faktor yang memengaruhi penilaian calon siswa. Hal ini akan membantu sekolah dalam meningkatkan kualitas proses seleksi siswa baru mereka dan memberikan pengalaman yang lebih baik kepada calon siswa dan orangtua mereka.

Dengan demikian, perkembangan teknologi informasi, terutama dalam hal text mining dan TF-IDF, dapat memberikan manfaat besar dalam meningkatkan efisiensi dan akurasi proses penerimaan siswa baru di sekolah Yayasan Perguruan. Hal ini mencerminkan bagaimana teknologi informasi terus berkembang dan dapat digunakan untuk mengatasi berbagai permasalahan di berbagai bidang kehidupan.

2. METODOLOGI PENELITIAN

2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) (pemrosesan/pengolahan bahasa alami) merupakan suatu metode yang mampu memproses input teks menjadi kata-kata kunci jawaban. Penemuan pengetahuan dapat menjadi efektif digunakan dan memperbaharui pola penemuan, lalu menerapkannya ke text mining. Text mining memiliki tahapan-tahapan yang dilakukan secara umum, antara lain: tokenizing, filtering, stemming, tagging, dan analyzing. Text mining merupakan suatu penemuan dari pengetahuan di dalam dokumen teks yang dapat mengungkapkan wawasan berharga dari teks-teks yang ada [6][7]. Pada tahapan pemrosesan pada suatu kata menggunakan Natural Language Processing dilakukan pada tiga tahap, yaitu: tokenizing, yang dapat digunakan sebagai pemecah kalimat menjadi kumpulan kata; filtering, digunakan untuk menghapus kata depan dan kata sambung dari suatu kumpulan kata yang masuk; dan untuk stemming digunakan untuk menghapus imbuhan. Penemuan pengetahuan dapat menjadi efektif digunakan dan memperbaharui pola penemuan, lalu menerapkannya ke text mining. Text mining merupakan suatu penemuan dari pengetahuan di dalam dokumen teks yang dapat mengungkapkan wawasan berharga dari teks-teks yang ada [8][9][10].

2.2 Metode Text Mining

Text mining merupakan suatu penemuan dari pengetahuan di dalam dokumen teks yang dapat mengungkapkan wawasan berharga dari teks-teks yang ada [11][12]. Text mining ini merupakan suatu tantangan dalam mendapatkan pengetahuan yang akurat pada dokumen teks dalam membantu pengguna menemukan apa yang diinginkan. Text mining merupakan suatu penemuan dari pengetahuan di dalam dokumen teks yang dapat mengungkapkan wawasan berharga dari teks-teks yang ada. [13]. Penemuan pengetahuan dapat menjadi efektif digunakan dan memperbaharui pola penemuan, lalu menerapkannya ke text mining. Text mining memiliki tahapan-tahapan yang dilakukan secara umum, antara lain: tokenizing, filtering, stemming, tagging, dan analyzing. Text mining merupakan suatu penemuan dari pengetahuan di dalam dokumen teks yang dapat mengungkapkan wawasan berharga dari teks-teks yang ada [11][14].

2.3 Algoritma Term Frequency Inverse Document Frequency (TF IDF)

Algoritma TF-IDF adalah suatu teknik untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval [15]. Pada teknik ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Algoritma Term Frequency Inverse Document Frequency (TFIDF) adalah suatu cara pemberian bobot hubungan suatu kata (term) terhadap dokumen. Term Frequency Inverse Document Frequency ini adalah sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen atau dalam sekelompok kata [11]. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata semakin besar jikasering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen. Di dalam Term Frequency Inverse Document Frequency untuk menghitung bobot (W) masing-masing dokumen yakni [16]:

$$Wdt = tfdt * Idft \quad (1)$$

$$W = Tf * (-) \quad (2)$$

Dimana:

Wdt : Bobot dokumen ke-d terhadap kata ke-t

tfdt : Banyaknya kata yang dicari pada sebuah

dokumenIdft : Inversed Document Frequency

$(\log(N/df))$

N : Total Dokumen

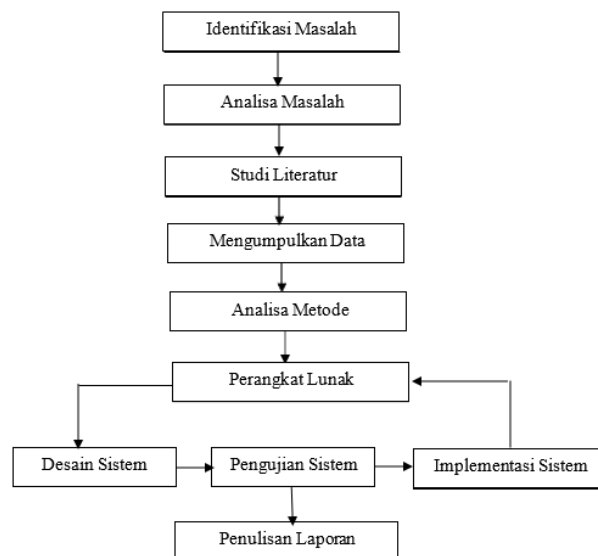
df : Banyaknya dokumen yang mengandung kata yang dicari

Berdasarkan rumus (2), seberapa pun besarnya nilai tf, apabila $N = df$ dimana sebuah kata/term di semua dokumen, maka akan didapatkan hasil 0 (nol) untuk perhitungan idf, sehingga perhitungan untuk bobot diubah menjadi:

$$W = Tf * \quad (3)$$

2.3 Kerangka Kerja Penelitian

Pada bab ini menjelaskan kerangka kerja dari penelitian yang akan dijelaskan secara rinci dengan alat bantu analisis dan bahan, materi dan urutan alur penelitian yang telah dibuat secara sistematis. Contoh kerangka penelitian dapat dilihat pada gambar 1 dibawah ini.



Gambar 1. Kerangka Kerja Penelitian

1. Identifikasi Masalah

Pada tahap identifikasi masalah digunakan untuk menjelaskan suatu masalah dan membuat penjelasan yang dapat diukur serta dapat digunakan sebagai petunjuk didalam penelitian.

2. Analisa Masalah

Pada tahap ini dilakukan analisa masalah serta memahami bidang masalah pada penelitian yang akan dikerjakan, menentukan tujuan perbaikan serta mengkomunikasikan penemuan-penemuan dan rekomendasi.

3. Studi Literatur

Pada studi literatur ini dilakukan dengan mengumpulkan dari berbagai pustaka yang nantinya akan dijadikan sebagai referensi penelitian.

4. Mengumpulkan Data

Pada tahap ini menggunakan data primer atau data sekunder. Untuk data primer didapatkan dari hasil wawancara, kuesioner dan observasi yang dilakukan dilapangan yang sesuai dengan penelitian ini. Untuk data sekunder didapatkan dari dokumen atau referensi yang ada berupa dokumen perencanaan jangka panjang atau pendek.

5. Analisa Metode

Analisa metode ini digunakan sebagai gambaran untuk mendeskripsikan metode yang digunakan sebagai solusi permasalahan yang ditimbulkan.

6. Perangkat Lunak

Dalam pembuatan perangkat lunak ini dibagi menjadi beberapa tahap yakni:

a. Desain Sistem

Pada tahap desain sistem ini merupakan suatu gambaran kepada user tentang sistem yang baru atau sebagai identifikasi komponen sistem yang akan didesain secara rinci.

- b. Pengujian Sistem
Pengujian sistem digunakan sebagai langkah untuk memastikan kualitas dari sistem yang telah dibangun dan memeriksa keselarasan antara komponen sistem yang akan diimplementasikan.
 - c. Implementasi Sistem
 - d. Pada tahapan ini menjelaskan mengenai manual modul ke semua user yang akan menggunakan sistem nantinya, sehingga user dapat merespon yang ditampilkan di sistem serta memberikan suatu masukan kepada pembuat sistem untuk dilakukan perbaikan agar sistem lebih baik lagi.
7. Penulisan Laporan
Pada tahap terakhir ini digunakan sebagai penulisan laporan data dan hasil yang diperoleh selama melakukan pengerjaan.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Selama ini pihak sekolah Yayasan Perguruan mengolah data penilaian calon siswa tersebut menggunakan Microsoft Excel, Namun cara tersebut belum bisa membuat pihak sekolah merasa puas karena dalam menentukan penilaian terhadap calon siswa tersebut membutuhkan waktu relatif lama, sehingga disaat menentukan kelulusan calon siswa memerlukan waktu selama satu minggu. Pihak sekolah merasa cara kerja sistem lama ini perlu diperbaiki, yang mana pihak sekolah menginginkan adanya suatu metode baru yang bisa mengatasi semua kendala dan permasalahan ini, agar proses penilaian terhadap calon siswa dapat diterima dengan cepat dari yang sebelumnya dan hasilnya maksimal mulai dari proses pemasukan data sampai dengan hasil kelulusan calon siswa. Hal ini penting diterapkan agar sistem penilaian calon siswa dan data siswa bisa menjadi suatu informasi yang bermanfaat dan dapat membantu pihak sekolah dalam menentukan kelulusan dari calon siswa. Untuk mengatasi permasalahan dalam penerimaan siswa baru tersebut maka dibutuhkan suatu metode yang cocok dengan menggunakan algoritma text mining dan TF- IDF. Algoritma text mining dan TF-IDF digunakan untuk membantu sistem dalam menjalankan sistem informasi penerimaan siswa baru dengan cara mengelompokkan data dari calon siswa baru.

3.2 Penerapan Algoritma Text Mining dan TF-IDF

Setiap tahunnya, SMP Yayasan Perguruan menerima ratusan data dari calon siswa yang akan mendaftarkan diri ke sekolah tersebut. Hal tersebut menjadi tuntutan sekolah agar lebih efisien dalam menerima dokumen yang diberikan oleh calon siswa. Data yang digunakan dalam penelitian ini adalah data yang berisikan berkas dan dokumen para calon siswa SMP Yayasan Perguruan yang berbentuk file HTML (HyperText Markup Language). Dokumen untuk pencarian dapat dilihat pada tabel 1.

Tabel 1. Dokumen Untuk Pencarian

Dokumen	Pertanyaan	Jawaban
D1	Berkas apa saja yang dibutuhkan untuk pendaftaran	Ijazah, KK, dll
D2	Apa ijazah asli dibutuhkan untuk penerimaan siswa baru	Ya, Tentu
D3	Sertifikat asli prestasi kejuaran diwajibkan dalam pendafran?	Tidak
D4	Apa berkas yang bersifat sertifikat wajib di legalisir?	Ya, Tentu
D5	Pas foto diwajibkan dalam penerimaan siswa baru?	Ya, Tentu
D6	Foto yang akan di upload harus dalam format apa	Jpg
D7	Semua berkas yang dibutuhkan harus diupload semua?	Ya, Tentu
D8	KK diwajibkan untuk pendaftaran?	Ya, Tentu
D9	KK dan berkas lainnya harus memiliki alamat yang sama?	Ya, Tentu
D10	Pas foto diwajibkan berwarna hitam putih?	Ya, Tentu
D11	Domisili diperlukan untuk berkas?	Ya, Tentu
D12	Surat vaksinasi termasuk dari berkas yang diperlukan atau tidak?	Ya, Tentu
D13	SKTM apa boleh diikutkan dalam pemberian berkas?	Ya, Tentu

Adapun langkah-langkah penggunaan metode text mining dan TF IDF adalah sebagai berikut :

1. Tokenizing

Tokenizing merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Dokumen akan melalui proses pemotongan menjadi perkata serta menghilangkan tanda baca yang ada sehingga tidak menghambat prosedur. Berikut ini merupakan hasil tokenizing dokumen dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Hasil *Tokenizing* Dokumen

No	Pertanyaan	Hasil <i>Tokenizing</i>
1	berkas apa saja yang dibutuhkan untuk pendaftaran	berkas-apa-saja-yang-dibutuhkan-untuk-pendaftaran
2	apa ijazah asli dibutuhkan untuk penerimaan siswa baru	apa-ijazah-asli-dibutuhkan-untuk-penerimaan-siswa-baru
3	Sertifikat asli prestasi kejuaran diwajibkan daalam pendaftaran	Sertifikat-asli-prestasi-kejuaran-diwajibkan-daalam-pendaftaran
4	Apa berkas yang bersifat sertifikat wajib di legalisir	Apa-berkas-yang-bersifat-sertifikat-wajib-di-legalisir
5	Pas foto diwajibkan dalam penerimaan siswa baru	Pas-foto-diwajibkan-dalam-penerimaan-siswa-baru
6	Foto yang akan di upload harus dalam format apa?	Foto-yang-akan-di-upload-harus-dalam-format-apa?
7	Semua berkas yang dibutuhkan harus di upload semua?	Semua-berkas-yang-dibutuhkan-harus-di-upload-semua?
8	KK diwajibkan untuk pendaftaran?	kk-diwajibkan-untuk-pendaftaran?
9	KK dan berkas lainnya harus memiliki alamat yang sama?	kk-dan-berkas-lainnya-harus-memiliki-alamat-yang-sama?
10	Pas foto diwajibkan berwarna hitam putih?	Pas-foto-diwajibkan-berwarna-hitam-putih?
11	Domisili diperlukan untuk berkas?	Domisili-diperlukan-untuk-berkas?
12	Surat vaksinasi termasuk dari berkas yang diperlukan atau tidak?	Surat-vaksinasi-termasuk-dari-berkas-yang-diperlukan-atau-tidak?
13	SKTM apa boleh diikutkan dalam pemberian berkas?	Sktm-apa-boleh-diikutkan-dalam-pemberian-berkas?

2. *Filtering*

Filtering merupakan tahap mengambil kata-kata penting dari hasil token yang dapat dilakukan dengan menggunakan stop list (membuang kata yang kurang penting) atau word list (menyimpan kata yang penting). Berikut merupakan hasil *filtering* dokumen dapat dilihat pada tabel 3 dibawah ini.

Tabel 3. Hasil *Filtering* Dokumen

No	Pertanyaan	Hasil <i>Tokenizing</i>
1	berkas- apa saja yang dibutuhkan- untuk pendaftaran	berkas-dibutuhkan-pendaftaran
2	apa ijazah-asli-dibutuhkan- untuk penerimaan-siswa- baru	ijazah-asli-dibutuhkan-penerimaan-siswa
3	sertifikat-asli-prestasi-kejuaran-diwajibkan- daalam -pendaftaran	sertifikat-asli-prestasi-kejuaran-diwajibkan-pendaftaran
4	apa berkas-yang-bersifat-sertifikat-wajib-di-legalisir	berkas-bersifat-sertifikat-wajib-legalisir
5	pas-foto-diwajibkan- dalam penerimaan-siswa- baru	pas-foto-diwajibkan-penerimaan-siswa
6	foto-yang akan di upload harus dalam format- apa	foto-upload-format
7	semua berkas-yang-dibutuhkan-harus-di-upload- semua	berkas-dibutuhkan-upload
8	kk-diwajibkan- untuk pendaftaran	kk-diwajibkan-pendaftaran
9	kk- dan berkas-lainnya-harus-memiliki-alamat-yang-sama	kk-berkas-lainnya-memiliki-alamat-yang-sama
10	pas-foto-diwajibkan-berwarna-hitam-putih	pas-foto-diwajibkan-berwarna-hitam-putih
11	domisili- diperlukan untuk-berkas	domisili-berkas
12	surat-vaksinasi-termasuk-dari-berkas-yang- diperlukan - atau - tidak	surat-vaksinasi-berkas
13	sktm- apa boleh diikutkan- dalam pemberian-berkas	sktm-diikutkan-pemberian-berkas

3. *Stemming*

Stemming merupakan tahap mencari *root* dari tiap kata hasil *filtering*. *Stemming* juga merupakan salah satu tahapan dalam pre processing dengan cara menghilangkan imbuhan yang terdapat pada kata. Berikut merupakan hasil *stemming* dokumen dapat dilihat pada tabel 4 dibawah ini.

Tabel 4. Hasil *Stemming* Dokumen

No	Pertanyaan	Hasil <i>Tokenizing</i>
1	berkas-dibutuhkan-pendaftaran	berkas- butuh-daftar
2	ijazah-asli-dibutuhkan-penerimaan-siswa	ijazah-asli- butuh -penerimaan-siswa
3	sertifikat-asli-prestasi-kejuaran-diwajibkan-pendaftaran	sertifikat-asli-prestasi-kejuaran- wajib-daftar
4	berkas-bersifat-sertifikat-wajib-legalisir	berkas- sifat -sertifikat-wajib-legalisir
5	pas-foto-diwajibkan-penerimaan-siswa	pas-foto- wajib -penerimaan-siswa
6	foto-upload-format	foto-upload-format
7	berkas-dibutuhkan-upload	berkas- butuh -upload
8	kk-diwajibkan-pendaftaran	kk- wajib-daftar
9	kk-berkas-lainnya-memiliki-alamat-yang-sama	kk-berkas-lainnya-memiliki-alamat

No	Pertanyaan	Hasil <i>Tokenizing</i>
10	pas-foto-diwajibkan-berwarna-hitam-putih	pas-foto- wajib-warna -hitam-putih
11	domisili-berkas	domisili-berkas
12	surat-vaksinasi-berkas	surat-vaksinasi-berkas
13	sktm-diikutkan-pemberian-berkas	sktm- ikut-beri -berkas

4. TF-IDF

Merupakan metode yang menggunakan dasar nilai statistik untuk memperlihatkan munculnya suatu kata yang terdapat pada dokumen. TF (term frequency) adalah frekuensi kemunculan term (t) pada dokumen. DF (document frequency) adalah banyaknya dokumen dimana suatu term (t) muncul. Perhitungan TF-IDF untuk mendapatkan bobot dari peroleh preprocessing disajikan pada tabel 5 di bawah ini.

Tabel 5. Menghitung TF-IDF

Term														df	D/df	IDF	IDF+1
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13				
Berkas	1			1			1		1		1	1	1	7	0.0458	1.3396	2.33959
Butuh	1	1					1							3	0.0196	1.7076	2.70757
Daftar	1		1					1	1					4	0.0261	1.5826	2.56823

Setelah memperoleh hasil TF-IDF, tahap selanjutnya adalah menghitung bobot seperti tabel 6 di bawah ini.

Tabel 6. Perhitungan Bobot W

Term (Q)	W = $tf * (IDF + 1)$												
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13
Berkas	2.34	0	0	2.34	0	0	2.34	0	2.34	0	2.34	2.34	2.34
Butuh	2.708	2.708	0	0	0	0	2.708	0	0	0	0	0	0
Daftar	2.583	0	2.583	0	0	0	0	2.583	2.583	0	0	0	0
Total	7.63	2.708	2.583	2.34	0	0	5.047	2.583	4.922	0	2.34	2.34	2.34
	5.474	0	0	5.474	0	0	Q	0	5.474	0	5.474	5.474	5.474
	7.331	7.331	0	0	0	0		0	0	0	0	0	0
	0	0	0	0	0	0		0	0	0	0	0	0
Total	12.8	7.331	0	5.474	0	0		0	5.474	0	5.474	5.474	5.474

3.3 Implementasi

1. Halaman utama

Pada halaman utama ini terdiri dari id, dan bantuan, ketika pengguna ingin menginputkan sebuah pertanyaan maka diharuskan untuk menginput id atau nama untuk sebuah tanda pengenalan kepada sistem chat bot ketika sudah melakukan penginputan identitas maka akan langsung masuk ke tampilan halaman percakapan, dan Tampilan halaman utama dapat dilihat pada gambar 2.



Gambar2. Tampilan Menu Utama

3.4 Hasil Pengujian

Dari hasil penerapan algoritma text mining dan algoritma TF-IDF pada analisa optimasi penerimaan siswa, maka penulis telah melakukan pengujian terhadap masing- masing pertanyaan, maka didapatkanlah hasil sebagai berikut:

Tabel 6. Perhitungan

No	Pertanyaan	Jawaban
1	apa ijazah asli dibutuhkan untuk penerimaan siswa baru	Iya, benar
2	Sertifikat asli prestasi kejuaran diwajibkan dalam pendaftaran	Tidak
3	Apa berkas yang bersifat sertifikat wajib di legalisir	Tidak
4	Pas foto diwajibkan dalam penerimaan siswa baru	Iya, benar
5	KK diwajibkan untuk pendaftaran?	Iya, benar
6	KK dan berkas lainnya harus memiliki alamat yang sama?	Tidak
7	Pas foto diwajibkan berwarna hitam putih?	Iya, benar
8	Domisili diperlukan untuk berkas?	Iya, benar

4. KESIMPULAN

Setelah menyelesaikan proses perancangan, implementasi, dan pengujian, dapat disimpulkan bahwa melalui hasil pemetaan pemodelan proses yang melibatkan tahapan preprocessing, penggunaan metode TF-IDF, dan perhitungan similarity, telah berhasil ditemukan tingkat kecocokan yang signifikan antara berbagai aspek pemodelan yang dilakukan. Kecocokan ini diidentifikasi berdasarkan nilai persamaan tertinggi yang diperoleh dari setiap aktivitas yang ada dalam proses tersebut. Selain itu, dalam tahap penyusunan komposisi web menggunakan algoritma text mining dan pemanfaatan metode TF-IDF, sistem berhasil beroperasi secara efektif dan efisien. Sistem ini mampu menghasilkan komposisi web secara otomatis tanpa gangguan yang berarti, memungkinkan pengguna untuk dengan cepat dan mudah menghasilkan konten web yang relevan dan berkualitas tinggi. Hasil ini menunjukkan bahwa pendekatan yang digunakan dalam pengembangan sistem ini telah menghasilkan hasil yang memuaskan dalam mengotomatiskan proses penyusunan konten web.

REFERENCES

- [1] L. Marlinah, "Pentingnya peran perguruan tinggi dalam mencetak SDM yang berjiwa inovator dan technopreneur menyongsong era society 5.0," *IKRAITH-EKONOMIKA*, vol. 2, no. 3, pp. 17–25, 2019.
- [2] R. Santoso, H. A. Munawi, and D. Sukmawati, "Perkembangan teknologi informasi dan telekomunikasi terhadap perubahan perilaku masyarakat," in *Prosiding 1st Corcys 2019, Conference on Research & Community Services*, P3M STKIP PGRI Jombang, 2019, pp. 586–592.
- [3] A. Taufik, "Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel," *J. Tek. Komput. AMIK BSI*, vol. 4, no. 2, pp. 112–118, 2018.
- [4] I. O. DjajaPutra, K. R. Priianti, and P. L. T. Irawan, "IMPLEMENTASI TEXT MINING UNTUK ANALISIS OPINI MASYARAKAT TERHADAP KINERJA LAYANAN TRANSPORTASI ONLINE DENGAN ANALISIS FAKTOR," *J. Simantec*, vol. 8, no. 2, pp. 45–53, 2020.
- [5] R. Melita, "Penerapan metode term frequency inverse document frequency (tf-IDF) dan cosine similarity pada sistem temu kembali INFORMASI untuk mengetahui Syarah Hadits Berbasis web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2018.
- [6] W. Nengsih, M. M. Zein, and N. Hayati, "Coarse-grained sentiment analysis berbasis natural language processing—ulasan hotel," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 1, pp. 41–48, 2021.
- [7] H. Herwin and K. Andesa, "Super agent chatbot '3s' sebagai media informasi menggunakan metoda natural language processing (NLP)," *J. Teknol. Dan Open Source*, vol. 2, no. 1, pp. 53–64, 2019.
- [8] R. Khoirunisa, "Penggunaan Natural Language Processing Pada Chatbot Untuk Media Informasi Pertanian," *IJAI (Indonesian J. Appl. Informatics)*, vol. 4, no. 2, pp. 55–63, 2020.
- [9] A. Y. Chandra, D. Kurniawan, and R. Musa, "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop)," *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 208–215, 2020.
- [10] K. Chowdhary and K. R. Chowdhary, "Natural language processing," *Fundam. Artif. Intell.*, pp. 603–649, 2020.
- [11] D. A. C. Rachman, R. Goejantoro, and F. D. T. Amijaya, "Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering," *EKSPONENSIAL*, vol. 11, no. 2, pp. 167–174, 2021.
- [12] F. F. Mailoa, "Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia," *J. Inf. Syst. Public Heal.*, vol. 6, no. 1, pp. 44–51, 2019.
- [13] A. Deolika, K. Kusri, and E. T. Luthfi, "Analisis pembobotan kata pada klasifikasi text mining," *J. Teknol. Inf.*, vol. 3, no. 2, pp. 179–184, 2019.
- [14] M. S. Hudin, M. A. Fauzi, and S. Adinugroho, "Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5518–5524, 2018.
- [15] A. Rokhim, "Implementasi Metode Term Frequency Inversed Document Frequence (Tf-Idf) dan Vector Space Model pada Aplikasi Pemberkasan Skripsi Berbasis WEB," *SPIRIT*, vol. 9, no. 1, 2018.
- [16] A. D. Sipayung, F. Fauziah, and N. Nurhayati, "Sistem Aplikasi Penilaian Jawaban Essay Test Calon Karyawan PT Siloam Hospitals TB Simatupang Menggunakan Algoritma Text Mining TF-IDF Berbasis Web," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 872–878, 2020.