

Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF

Natalia Silalahi^{1*}, Guidio Leonarde Ginting²

¹Fakultas Ilmu Pendidikan, PGSD, Universitas Negeri Medan, Medan, Indonesia

²Fakultas Ilmu Komputer dan Teknologi Informasi, Sistem Informasi, Universitas Budi Darma, Medan, Indonesia

Email: ^{1*}natalia.novena.silalahi@gmail.com, ²guidio.leonard626@gmail.com

Email Penulis Korespondensi: natalia.novena.silalahi@gmail.com

Abstrak—Penyajian berita pada umumnya disusun dengan sedemikian rupa agar informasi yang disajikan terkelompok dengan baik, namun penggunaan media elektronik tidak serta merta dapat menawarkan kategori berita dengan lengkap dikarenakan space yang ditawarkan tidak semua dapat dipenuhi dengan penyajian yang baik, sehingga diperlukan treatment khusus agar pembaca mendapatkan berita yang dibutuhkan yang tersusun berdasarkan rekomendasi. Untuk menyusun penelitian ini agar lebih terstruktur, maka penulis melakukan beberapa tahapan dalam menyelesaikan penelitian tersebut yaitu Tahap Identifikasi Masalah, Tahap Studi Literatur, Tahap Pengumpulan Data, Tahap Implementasi Algoritma Text mining dan TF-IDF, dan kesimpulan. Penulis melakukan implementasi algoritma text mining dan TF-IDF dalam mengolah data judul berita diawali dengan Algoritma Text Mining dimana tahapan tersebut sebagai tahapan preprosesing dengan tujuan agar data yang akan diolah merupakan kata dasar sehingga proses pembobotan pada Algoritma TF-IDF tidak terlalu luas. Setelah tahapan text mining maka akan dilanjutkan ke tahapan TF-IDF yaitu pembobotan terhadap term pada setiap dokumen. Algoritma text mining dan TF-IDF mampu memberikan rekomendasi berita sesuai berdasarkan kemiripan makna tertinggi baik secara topik maupun objek dari judul berita, untuk penelitian berikutnya disarankan untuk menggunakan algoritma lain seperti cosine similiary agar rekomendasi tidak hanya dihasilkan dari kesesuaian kata namun dapat juga melihat kesamaan makna sehingga hasil penelitian dapat lebih baik lagi

Kata Kunci: Rekomendasi; Berita; Terkait; Text Mining; TF-IDF; NLP

Abstract—News presentation is generally structured in such a way that the information presented is well grouped, but the use of electronic media does not necessarily offer complete news categories because not all of the space offered can be filled with good presentation, so special treatment is needed so that readers get the news. needed which is arranged based on recommendations. To arrange this research to be more structured, the authors carried out several stages in completing the research, namely the Problem Identification Stage, Literature Study Stage, Data Collection Stage, Text Mining and TF-IDF Algorithm Implementation Stage, and conclusions. The author implements the text mining and TF-IDF algorithms in processing news title data starting with the Text Mining Algorithm where this stage is a preprocessing stage with the aim that the data to be processed is a basic word so that the weighting process in the TF-IDF Algorithm is not too broad. After the text mining stage, it will proceed to the TF-IDF stage, namely weighting the terms in each document. Text mining and TF-IDF algorithms are able to provide appropriate news recommendations based on the highest similarity in meaning both in terms of topic and object of the news title, for future research it is recommended to use other algorithms such as cosine similar so that recommendations are not only generated from the suitability of words but can also see the similarity of meaning so that research results can be even better.

Keywords: Recommendation; News; Related; Text Mining; TF-IDF, NLP

1. PENDAHULUAN

Teknologi informasi sangat berkembang pesat, perkembangan tersebut dapat terlihat dari kemudahan mendapatkan informasi, berkomunikasi sampai dengan melakukan transaksi keuangan, kemudahan tersebut telah mempengaruhi perilaku masyarakat dalam melakukan kegiatan sehari-hari seperti sosialisasi, perdagangan, kesehatan, pendidikan, rekreasi, dan masih banyak lainnya [1][2]. Perkembangan teknologi informasi telah menciptakan hal baru yang mempengaruhi berbagai aspek kehidupan, seperti kesehatan, pendidikan, perdagangan, perbankan, dan lainnya, perubahan tersebut dapat terlihat dari layanan yang diberikan telah berbasis elektronik yang semua diawali dengan “e” seperti e-library, e-learning, e-commerce, e-banking, dan lainnya [3]. Transformasi teknologi informasi yang dulu dianggap canggih namun teknologi saat ini jauh lebih baik lagi, hal tersebut terjadi karena kebutuhan masyarakat terhadap teknologi informasi dalam menyelesaikan masalah dituntut untuk dapat lebih baik lagi.

Berita merupakan informasi terhadap suatu peristiwa yang telah terjadi dan disebar luaskan melalui media massa cetak atau media massa elektronik. Informasi merupakan hasil yang diperoleh dari pengolahan data yang dapat berupa kabar, keterangan atau pemberitahuan [4]. Peran teknologi informasi dalam penyebaran informasi berita sangat signifikan, perubahan tersebut dapat terlihat dari pemanfaatan media elektronik sebagai media penyebaran informasi berita yang dapat diakses dimana saja dan kapan saja, kemudian tidak hanya itu perubahan lainnya ialah manfaat teknologi informasi yang ditawarkan seperti kecepatan dalam mendapatkan informasi, unsur berita yang lebih lengkap, seperti audio, video bahkan fasilitas diskusi. Penerapan teknologi informasi berita juga tidak terlepas dari teknologi analisa berita yang semakin baik untuk memenuhi kebutuhan dari masyarakat terhadap informasi berita.

Rekomendasi merupakan sesuatu bentuk komunikasi sekaligus promosi secara tidak langsung sebagai ajakan untuk melakukan hal lain yang berkaitan dengan suatu hal yang telah digunakan [5]. Rekomendasi berita dapat diartikan sebagai fitur pada media massa elektronik yang menawarkan berita lain yang berkaitan terhadap berita yang sedang dibaca. Penyajian berita pada umumnya disusun dengan sedemikian rupa agar informasi yang disajikan terkelompok dengan baik, namun penggunaan media elektronik tidak serta merta dapat menawarkan kategori berita dengan lengkap dikarenakan space yang ditawarkan tidak semua dapat dipenuhi dengan penyajian yang baik, sehingga diperlukan treatment

husus agar pembaca mendapatkan berita yang dibutuhkan yang tersusun berdasarkan rekomendasi. Rekomendasi akan membantu pembaca untuk tetap dapat mengikuti berita yang berkaitan terhadap berita yang sedang atau telah dibaca yang tersedia pada media berita tersebut.

Algoritma Text Mining Text mining dapat didefinisikan secara luas sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining yang salah satunya adalah kategorisasi. Text mining merupakan proses menambang text yang tidak terstruktur menjadi format yang terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru [6]. Pada penelitian yang pertama text mining classification sebagai rekomendasi dosen pembimbing tugas akhir berhasil dilakukan dengan akurasi 60% dengan menggunakan 98 data training dan 10 data testing [7]. Penelitian kedua dengan menggunakan algoritma text mining untuk mengklasifikasi judul berita online studi kasus radar banjar masin menggunakan metode naïve bayes, berdasarkan analisis yang dilakukan penulis metode tersebut mampu mengklasifikasikan berita online dengan menggunakan 400 data dan 4 kategori dapat menghasilkan akurasi sebanyak 78,75% sedangkan hasil recall adalah 80, 56% dan precision adalah 78,75% [8]. Pada penelitian selanjutnya sistem rekomendasi dari berita Microsoft news dengan menggunakan algoritma TF-IDF dan Cosine Similarity dapat merekomendasikan 10 berita terbaik dengan menggunakan data set 5.286 riwayat pembaca yang setidaknya mengklik 40 berita dan 5000 judul berita, penelitian tersebut menghasilkan nilai akurasi sebesar 80,77% [9].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk menyusun penelitian ini agar lebih terstruktur, maka penulis melakukan beberapa tahapan dalam menyelesaikan penelitian tersebut.

1. Tahap Identifikasi Masalah

Pada tahap ini, penulis menguraikan apa yang menjadi sumber masalah dalam layanan customer service RSUD Pirngadi Kota Medan.

2. Tahap Studi Literatur

Pada tahap ini, penulis mempelajari dan memahami teori – teori yang menjadi pedoman dan referensi yang diperoleh dari berbagai buku, jurnal, situs yang relevan terkait chatbot, implementasi algoritma atau metode yang digunakan, serta teori – teori yang menunjang materi penelitian.

3. Tahap Pengumpulan Data

Pengumpulan data adalah suatu metode yang digunakan untuk mendapatkan suatu informasi untuk kebutuhan penelitian. Data yang digunakan yaitu data hasil survei dan wawancara kepada pihak RSUD Pirngadi Kota Medan mengenai layanan customer service RSUD Pirngadi Kota Medan.

4. Tahap Implementasi Algoritma Text mining dan TF-IDF

Pada tahap tersebut, penulis melakukan implementasi algoritma text mining dan TF-IDF dalam mengolah data judul berita, penerapan algoritma diawali dengan Algoritma Text Mining dimana tahapan tersebut sebagai tahapan preprocessing dengan tujuan agar data yang akan diolah merupakan kata dasar sehingga proses pembobotan pada Algoritma TF-IDF tidak terlalu luas. Setelah tahapan text mining maka akan dilanjutkan ke tahapan TF-IDF yaitu pembobotan terhadap term pada setiap dokumen.

5. Kesimpulan

Sebagai tahap akhir penelitian tersebut maka penulis akan menarik kesimpulan yang memuat hasil dari penelitian rekomendasi berita terkait menggunakan algoritma text mining dan TF – IDF dan hal – hal yang perlu diperbaiki.

2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan proses pengolahan bahasa natural dimana bahasa natural merupakan bahasa yang biasanya digunakan oleh manusia untuk berkomunikasi dalam kehidupan sehari-hari [10]. Natural Language Processing (NLP) juga merupakan bidang penelitian yang mengeksplorasi bagaimana komputer dapat memahami dan memanipulasi bahasa natural, baik itu dalam teks ataupun perkataan [11]. Natural Language Processing (NLP) digunakan untuk memperoleh pemahaman tentang bahasa manusia yang dilakukan oleh mesin, sehingga dapat berinteraksi dengan manusia dengan cara yang lebih alami. Untuk mengubah bahasa manusia menjadi data yang dapat diproses oleh komputer, NLP menggunakan algoritma dan metode khusus. Mesin belajar untuk memahami konteks, arti, dan tujuan dari sebuah kalimat atau teks [12]

2.3 Algoritma Text Mining

Text mining bisa didefinisikan sebagai knowledge-intensive process dimana pengguna berinteraksi dengan kumpulan dokumen seiring berjalannya waktu dengan menggunakan seperangkat alat analisis [13]. Text mining digunakan untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola-pola menarik [14]. Sumber data pada text mining adalah kumpulan dokumen yang berupa data teks yang tidak terstruktur. Dalam penelitian ini text mining digunakan untuk mendapatkan text summarization dari sumber data [15]. Secara umum tahapan- tahapan dari text preprocessing adalah sebagai berikut.

1. Case Folding: Tahap case folding adalah mengubah seluruh huruf dari „a“ sampai dengan „z“ dalam dokumen menjadi huruf kecil. Tidak semua dokumen konsisten dengan penggunaan huruf kapital. Maka dari itu case folding mengkonversi keseluruhan teks dalam dokumen menjadi huruf kecil.
2. Tokenization: tahap membagi data mentah menjadi unit-unit kecil yang disebut token atau kata. Tujuan dari tahap ini adalah untuk memecah data mentah menjadi bagian-bagian yang lebih kecil dan lebih mudah untuk dianalisis.
3. Filtering: Tahap filtering adalah tahap mengambil kata-kata penting dari hasil tokenizing menggunakan algoritma stopword removal (membuang kata yang kurang penting). Stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words.
4. Stemming: tahap mengubah kata-kata dalam data mentah menjadi bentuk dasar atau kata dasar. Tujuan dari tahap ini adalah untuk menghilangkan berbagai variasi kata yang sebenarnya memiliki makna yang sama [16].

2.4 Algoritma Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu proses untuk melakukan transformasi data dari data tekstual ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur [17]. Metode ini menghitung frekuensi kemunculan kata dalam dokumen (term frequency) dan memperhitungkan seberapa sering kata tersebut muncul di seluruh dokumen teks (inverse document frequency). Semakin sering kata muncul dalam dokumen tertentu, semakin tinggi skor term frequency nya, apabila semakin jarang muncul maka semakin tinggi skor inverse document frequency nya. Dengan menggunakan metode TF-IDF, kita dapat menemukan kata-kata yang paling penting dari suatu dokumen teks [15]. Pada Term Frequency (TF) terdapat beberapa jenis formula yang dapat digunakan [18], yaitu:

1. TF biner (binary TF), hanya memperlihatkan apakah suatu kata atau term ada atau tidak dalam dokumen, jika ada maka diberi nilai satu (1), jika tidak maka diberikan nilai nol (0).
2. TF murni (raw TF) merupakan nilai berdasarkan jumlah kemunculan suatu term di dokumen. Contohnya, jika muncul lima (5) kali maka kata diberikan nilai lima (5).
3. TF normalisasi, menggunakan perbandingan antar frekuensi sebuah term dengan nilai maksimum dari keseluruhan atau kumpulan frekuensi term yang ada pada dokumen.
4. TF logaritmik, hal ini menghindari dominansi dokumen yang mengandung sedikit term dalam query, namun mempunyai frekuensi yang tinggi.

Dalam penelitian ini, penulis menggunakan jenis formula untuk perhitungan adalah TF murni (raw TF) selain itu formula ini paling sering digunakan. Cara kerja dalam mencari nilai term-frequency melalui beberapa persamaan yaitu:

$$Tf_{t,d} = 1 + \log tf \quad (1)$$

keterangan dari variabel sebagai berikut:

tf : term frekuensi atau banyaknya kata pada dokumen

$Tf_{t,d}$: term frekuensi atau banyaknya kata t pada dokumen d atau pembobotan lokal

Mencari nilai inverse document-frequency melalui persamaan 2:

$$Idf_t = \log n/df_t \quad (2)$$

keterangan dari variabel sebagai berikut:

Idf_t : inverse document-frequency atau pembobotan global

n : banyaknya dokumen

df_t : banyaknya dokumen yang memiliki kata t .

Berdasarkan persamaan 1 dan persamaan 2 baru dapat ditentukan nilai bobotnya ($W_{t,d}$) dengan mengalikan kedua persamaan sehingga menjadi persamaan 3.

$$W_{t,d} = tf_{t,d} \cdot idf_t \quad (3)$$

keterangan:

$Tf_{t,d}$: term frekuensi atau banyaknya kata pada dokumen atau pembobotan lokal

idf_t : inverse document frequency atau pembobotan global

$W_{t,d}$: nilai bobot akhir kata

Pembobotan tf-idf digunakan untuk mengukur seberapa penting suatu kata dalam suatu dokumen.

3. HASIL DAN PEMBAHASAN

Pada pembahasan penelitian tersebut akan menjelaskan tahapan penerapan algoritma text mining dan Algoritma TF-IDF untuk mendapatkan hasil rekomendasi keterkaitan berita. Penulis membuat skenario rekomendasi berita berdasarkan judul berita yang memiliki keterkaitan yang paling tinggi, sebagai contoh berita awal yang dijadikan sebagai data pembanding terhadap data judul berita adalah “Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisinya Sendiri” kemudian data judul berita akan dibobotkan berdasarkan judul berita tersebut, hasil pembobotan tertinggi adalah judul berita yang terkait, pada penelitian tersebut penulis menggunakan data judul berita sebanyak 21 yang dapat dilihat pada tabel 1.

Tabel 1. Data judul berita

No	Judul berita
1.	Menanti Cawapres Anies yang Sudah 'Dikunci' NasDem-PD-PKS
2.	Edan! Tiket Formula E Jakarta Harga Rp 20 Juta Ludes Terjual
3.	Tiket Formula E Jakarta 2023 Diklaim Laku Keras
4.	Cara Beli Tiket Formula E Jakarta 2023, Lengkap dengan Lokasi Penukaran
5.	Biar Pebalap Tak Kepanasan, Formula E Jakarta 2024 Bakal Digelar Malam Hari
6.	Jadwal Formula E Jakarta 2023, Balapan Digelar Dua Kali
7.	Daftar Harga Tiket Formula E Jakarta, Termahal Tembus Rp 20 Juta!
8.	Formula E Jakarta 2023 Digelar Tanpa APBD dan Commitment Fee
9.	Warisan Era Anies, Ini Jadwal Balapan Mobil Listrik Formula E di Jakarta Tahun Ini
10.	Formula E Jakarta 2023 Pakai Ban Khusus, Ini Spesifikasinya
11.	Anies Ungkap Isi Pertemuan dengan SBY di Pacitan: Diskusi Politik
12.	Teka-teki Sook Bakal Cawapres Anies Baswedan Hasil Keputusan Tim 8
13.	Jawaban Koalisi Anies Balas Wanti-wanti Mahfud soal 'Dijegal Internal'
14.	Anies soal 1 Nama Cawapres: Proses Sudah Berjalan, Nanti Diumumkan
15.	NasDem Jatim Sebut AHY Pantas Jadi Cawapres Anies, Tapi...
16.	SBY Ajak AHY Temui Anies Baswedan di Pacitan, Sodorkan Jadi Cawapres?
17.	Empat Mata SBY-Anies Bahas Nasib AHY hingga Sentil Cawe-cawe Istana
18.	Blak-blakan Elite Demokrat SBY Sodorkan AHY Jadi Cawapres Anies
19.	3,5 Jam Pertemuan SBY-Anies di Pacitan, Elite Demokrat Bungkam
20.	Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisinya Sendiri
21.	Mahfud harap Anies dapat tiket pilpres jangan sampai dijegal koalisinya sendiri

3.1 Penerapan Algoritma Text mining

Adapun tujuan dari penerapan algoritma text mining pada penelitian tersebut ialah sebagai preprosesing dimana judul berita akan dipersiapkan dengan sedemikian rupa agar data tersebut dipastikan hanya kata yang memiliki arti sehingga proses perbandingan lebih sederhana. Sebagai contoh penerapan algoritma text mining terhadap data tersebut maka penulis mengambil salah satu data judul berita yaitu “Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisinya Sendiri” kemudian diolah sesuai dengan tahapan dari algoritma tersebut yang dapat dilihat pada Tabel 2.

Tabel 2. Penerapan Text Mining

Tahapan Text Mining	Tahapan Text Mining
Hasil Case Folding	mahfud harap anies dapat tiket pilpres jangan sampai dijegal koalisinya sendiri
Hasil Tokenizing	[mahfud] [harap] [anies] [dapat] [tiket] [pilpres] [jangan] [sampai] [dijegal] [koalisinya] [sendiri]
Hasil Filtering	[mahfud] [harap] [anies] [dapat] [tiket] [pilpres] [dijegal] [koalisinya]
Hasil Stemming	[mahfud] [harap] [anies] [dapat] [tiket] [pilpres] [jegal] [koalisi]

Berdasarkan tahapan diatas, maka hasil penerapan algoritma text mining terhadap seluruh data adalah sebagai berikut :

Tabel 3. Hasil Penerapan text mining

No	Judul berita	Hasil Text Mining
1.	Menanti Cawapres Anies yang Sudah 'Dikunci' NasDem-PD-PKS	[menanti] [cawapres] [anies] [dikunci] [nasdem] [pd] [pks]
2.	Edan! Tiket Formula E Jakarta Harga Rp 20 Juta Ludes Terjual	[edan] [tiket] [formula] [e] [jakarta] [harga] [rp] [20] [juta] [ludes] [terjual]
3.	Tiket Formula E Jakarta 2023 Diklaim Laku Keras	[tiket] [formula] [e] [jakarta] [2023] [diklaim] [laku] [keras]
4.	Cara Beli Tiket Formula E Jakarta 2023, Lengkap dengan Lokasi Penukaran	[beli] [tiket] [formula] [e] [jakarta] [2023] [lengkap] [lokasi] [penukaran]
5.	Biar Pebalap Tak Kepanasan, Formula E Jakarta 2024 Bakal Digelar Malam Hari	[biar] [pebalap] [kepanasan] [formula] [e] [jakarta] [2024] [digelar] [malam]
6.	Jadwal Formula E Jakarta 2023, Balapan Digelar Dua Kali	[jadwal] [formula] [e] [jakarta] [2023] [balapan] [digelar] [kali]
7.	Daftar Harga Tiket Formula E Jakarta, Termahal Tembus Rp 20 Juta!	[daftar] [harga] [tiket] [formula] [e] [jakarta] [termahal] [tembus] [rp] [20] [juta]
8.	Formula E Jakarta 2023 Digelar Tanpa APBD dan Commitment Fee	[formula] [e] [jakarta] [2023] [digelar] [apbd] [commitment] [fee]
9.	Warisan Era Anies, Ini Jadwal Balapan Mobil Listrik Formula E di Jakarta Tahun Ini	[warisan] [era] [anies] [jadwal] [balapan] [mobil] [listrik] [formula] [e] [jakarta]

No	Judul berita	Hasil Text Mining
10.	Formula E Jakarta 2023 Pakai Ban Khusus, Ini Spesifikasinya	[formula] [e] [jakarta] [2023] [pakai] [ban] [spesifikasinya]
11.	Anies Ungkap Isi Pertemuan dengan SBY di Pacitan: Diskusi Politik	[anies] [ungkap] [isi] [pertemuan] [sby] [pacitan] [diskusi] [politik]
12.	Teka-teki Sook Bakal Cawapres Anies Baswedan Hasil Keputusan Tim 8	[teka] [teki] [sook] [cawapres] [anies] [baswedan] [hasil] [keputusan] [tim] [8]
13.	Jawaban Koalisi Anies Balas Wanti-wanti Mahfud soal 'Dijegal Internal'	[koalisi] [anies] [balas] [wanti] [wanti] [mahfud] [soal] [dijegal] [internal]
14.	Anies soal 1 Nama Cawapres: Proses Sudah Berjalan, Nanti Diumumkan	[anies] [1] [nama] [cawapres] [proses] [berjalan] [diumumkan]
15.	NasDem Jatim Sebut AHY Pantas Jadi Cawapres Anies, Tapi...	[nasdem] [jatim] [ahy] [cawapres] [anies]
16.	SBY Ajak AHY Temui Anies Baswedan di Pacitan, Sodorkan Jadi Cawapres?	[sby] [ajak] [ahy] [temui] [anies] [baswedan] [pacitan] [sodorkan] [cawapres]
17.	Empat Mata SBY-Anies Bahas Nasib AHY hingga Sentil Cawe-cawe Istana	[mata] [sby] [anies] [bahas] [nasib] [ahy] [sentil] [cawe] [cawe] [istana]
18.	Blak-blakan Elite Demokrat SBY Sodorkan AHY Jadi Cawapres Anies	[blak] [blakan] [elite] [democrat] [sby] [sodorkan] [ahy] [jadi] [cawapres] [anies]
19.	3,5 Jam Pertemuan SBY-Anies di Pacitan, Elite Demokrat Bungkam	[3,5] [jam] [pertemuan] [sby] [anies] [pacitan] [elite] [demokrat] [bungkam]
20.	Museum SBY-ANI Saksi Bisu Pertemuan SBY dengan Dua Calon Presiden	[museum] [sby] [ani] [saksi] [bisu] [pertemuan] [sby] [calon] [presiden]
21.	Mahfud harap Anies dapat tiket pilpres jangan sampai dijegal koalisinya sendiri	[mahfud] [harap] [anies] [dapat] [tiket] [pilpres] [jegal] [koalisi]

3.2 Penerapan Algoritma TF-IDF

Sesuai dengan skenario yang telah disampaikan maka pada tahapan penerapan algoritma TF-IDF penulis menentukan term sebagai berita awal yang dijadikan data pembandingan terhadap data judul berita, Adapun tahapan dari algoritma tersebut sebagai berikut :

Tabel 4. Hasil Penerapan Algoritma TF-IDF

term	D1	D2	D3	D4	D5	D6	D7	D8	D9	...	D18	D19	D20	df	D/df	IDF
mahfud	0	0	0	0	0	0	0	0	0	..	0	0	0	1	20	1,301
harap	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0,000
anies	1	0	0	0	0	0	0	0	1	..	1	1	0	11	1,82	0,260
tiket	0	1	1	1	0	0	1	0	0	..	0	0	0	4	5	0,699
pilpres	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0,000
jegal	0	0	0	0	0	0	0	0	0	..	0	0	0	1	20	1,301
koalisi	0	0	0	0	0	0	0	0	0	..	0	0	0	1	20	1,301

Tabel 5. Tabel hasil pembobotan algoritma TF-IDF

term	D1	D2	D3	D4	D5	D6	D7	D8	D9	...	D14	D15	D16	D17	D18	D19	D20
mahfud	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0
harap	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0
anies	0,3	0	0	0	0	0	0	0	0,3	..	0,3	0,3	0,3	0,3	0,3	0,3	0
tiket	0	0,7	0,7	0,7	0	0	0,7	0	0	..	0	0	0	0	0	0	0
pilpres	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0
jegal	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0
koalisi	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0
Total	0,3	0,7	0,7	0,7	0	0	0,7	0	0,3	..	0,3	0,3	0,3	0,3	0,3	0,3	0

Berdasarkan hasil perhitungan algoritma TF-IDF maka didapatkan hasil pembobotan data judul berita terhadap data judul berita pembandingan “Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisi Sendiri” dapat dilihat pada tabel 5, berdasarkan hasil dari pembobotan tersebut setelah diurutkan dari bobot tertinggi dapat dilihat pada tabel 6, maka judul berita yang akan dijadikan sebagai rekomendasi berita selanjutnya adalah dari judul berita D13 yaitu “Menanti Cawapres Anies yang Sudah 'Dikunci' NasDem-PD-PKS”.

Tabel 6. Tabel rangking pembobotan Algoritma TF-IDF

Ranking	Dokumen	Judul Berita	Bobot
1	D13	Jawaban Koalisi Anies Balas Wanti-wanti Mahfud soal 'Dijegal Internal'	2,86
2	D12	Teka-teki Sook Bakal Cawapres Anies Baswedan Hasil Keputusan Tim 8	1,56

Ranking	Dokumen	Judul Berita	Bobot
3	D2	Edan! Tiket Formula E Jakarta Harga Rp 20 Juta Ludes Terjual	0,70
4	D3	Tiket Formula E Jakarta 2023 Diklaim Laku Keras	0,70
5	D4	Cara Beli Tiket Formula E Jakarta 2023, Lengkap dengan Lokasi Penukaran	0,70
6	D7	Daftar Harga Tiket Formula E Jakarta, Termahal Tembus Rp 20 Juta!	0,70
7	D1	Menanti Cawapres Anies yang Sudah 'Dikunci' NasDem-PD-PKS	0,26
8	D9	Warisan Era Anies, Ini Jadwal Balapan Mobil Listrik Formula E di Jakarta Tahun Ini	0,26
9	D11	Anies Ungkap Isi Pertemuan dengan SBY di Pacitan: Diskusi Politik	0,26
10	D14	Anies soal 1 Nama Cawapres: Proses Sudah Berjalan, Nanti Diumumkan	0,26
11	D15	NasDem Jatim Sebut AHY Pantas Jadi Cawapres Anies, Tapi...	0,26
12	D16	SBY Ajak AHY Temui Anies Baswedan di Pacitan, Sodorkan Jadi Cawapres?	0,26
13	D17	Empat Mata SBY-Anies Bahas Nasib AHY hingga Sentil Cawe-cawe Istana	0,26
14	D18	Blak-blakan Elite Demokrat SBY Sodorkan AHY Jadi Cawapres Anies	0,26
15	D19	3,5 Jam Pertemuan SBY-Anies di Pacitan, Elite Demokrat Bungkam	0,26
16	D5	Biar Pebalap Tak Kepanasan, Formula E Jakarta 2024 Bakal Digelar Malam Hari	0,00
17	D6	Jadwal Formula E Jakarta 2023, Balapan Digelar Dua Kali	0,00
18	D8	Formula E Jakarta 2023 Digelar Tanpa APBD dan Commitment Fee	0,00
19	D10	Formula E Jakarta 2023 Pakai Ban Khusus, Ini Spesifikasinya	0,00
20	D20	Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisinya Sendiri	0,00

4. KESIMPULAN

Kesimpulan dari hasil penelitian penerapan algoritma text mining dan TF-IDF dalam rekomendasi berita yang dilakukan penulis ialah algoritma tersebut mampu memberikan rekomendasi berita yang sesuai dengan judul berita data pebanding yaitu “Mahfud Harap Anies Dapat Tiket Pilpres, Jangan Sampai Dijegal Koalisinya Sendiri” yang menghasilkan judul berita rekomendasi “Menanti Cawapres Anies yang Sudah 'Dikunci' NasDem-PD-PKS” hasil rekomendasi judul berita tersebut memiliki kesesuaian secara objek maupun secara topik. Untuk penelitian berikutnya penulis menyarankan untuk menggunakan algoritma lain seperti cosine similarity agar rekomendasi tidak hanya dihasilkan dari kesesuaian kata namun dapat juga melihat kesamaan makna sehingga hasil penelitian dapat lebih baik lagi.

REFERENCES

- [1] D. Wiryany, S. Natasha, R. Kurniawan, J. I. Komunikasi, and M. Bandung, “PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI TERHADAP PERUBAHAN SISTEM KOMUNIKASI INDONESIA,” 2022.
- [2] J. Pendidikan and D. Konseling, “Perkembangan Teknologi Informasi dan Komunikasi (TIK) Terhadap Kulaitas Pembelajaran Di Sekolah Dasar Ikrham Abdaul Huda.”
- [3] Ardy Wirasaputra, Fikri Riduan, Pramudhya, Riyan, Zulkahfi, and S. Pd. , M. P. Widyah Noviana, “DAMPAK DARI PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI,” Jurnal Kreativitas Mahasiswa Informatika (JATIMIKA), vol. 3, no. 2, pp. 206–210, 2022.
- [4] F. Delfariyadi, A. Helen, and S. Yuliawati, “Klasifikasi Sentimen Judul Berita Pemberitaan COVID-19 Tahun 2021 pada Media DetikHealth.”
- [5] G. Ferio, R. Intan, and S. Rostianingsih, “Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity.”
- [6] IBM, “Text mining,” <https://www.ibm.com/topics/text-mining>.
- [7] Z. Efendi and M. Mustakim, “Text Mining Classification sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi,” Seminar Nasional Teknologi Informasi Komunikasi dan Industri, vol. 0, no. 0, pp. 235–242, 2017.
- [8] M. Sholih 'afif, M. Muzakir, M. I. Al, and G. Al Awalaen, “TEXT MINING UNTUK MENGGKLASIFIKASI JUDUL BERITA ONLINE STUDI KASUS RADAR BANJARMASIN MENGGUNAKAN METODE NAÏVE BAYES,” Kumpulan jurnal Ilmu Komputer (KLIK), vol. 08, no. 2, 2021.
- [9] G. Yunanda, D. Nurjanah, and S. Meliana, “Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods,” Building of Informatics, Technology and Science (BITS), vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1670.
- [10] R. Loukanova, Ed., Natural Language Processing in Artificial Intelligence — NLPinAI 2021, vol. 999. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-030-90138-7.
- [11] K. Aditama, “PEMANFAATAN NATURAL LANGUAGE PROCESSING DAN PATTERN MATCHING DALAM PEMBELAJARAN MELALUI GURU VIRTUAL,” ELKOM, vol. 13, no. 1, pp. 121–133, 2020, [Online]. Available: <http://ejurnal.stekom.ac.id/index.php/home□page121>
- [12] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
- [13] A. K. OJO and A. B. Adeyomo, “Framework for Knowledge Discovery from Journal Articles Using Text Mining Techniques,” African Journal Of Computing & ICT, vol. 5, no. 2, pp. 35–44, Mar. 2012.
- [14] D. Ariyanti and K. Iswardani, “Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes.”

- [15] Fikri Aldi Nugraha, Nisa Hanum Harani, and Roni Habibi, Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning. Bandung: Kreatif Industri Nusantara, 2020.
- [16] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, “Stemming Indonesian,” ACM Transactions on Asian Language Information Processing, vol. 6, no. 4, pp. 1–33, Dec. 2007, doi: 10.1145/1316457.1316459.
- [17] Y. S. Hartini, A. B. P. Lefanska, A. A. Ursia, D. A. B. Prasetyo, and B. Sugiharto, Prosiding Seminar Nasional Sanata Dharma Berbagi “Pengembangan, Penerapan Dan Pendidikan ‘Sains Dan Teknologi’ Pasca Pandemi.” yogyakarta: Sanata Dharma University Press, 2022.
- [18] M. R. Faisal, “Belajar Data Science: Text Mining Untuk Pemula I DNA Sequence Classification View project IT Asset Management View project,” 2022.