

Pengembangan Metode Information Retrieval dan Haversine Formula untuk Rekomendasi Penentuan Klinik di Kabupaten Jember

Fadhel Akhmad Hizham¹, Raden Venantius Hari Ginardi²

Program Studi Informatika, Institut Teknologi dan Bisnis Widya Gama Lumajang¹

Program Studi Informatika, Institut Teknologi Sepuluh November²

Corresponding Author: Marita Qori'atunnadyah (maritaqori@gmail.com)

ARTICLE INFO

Date of entry:

1 Oktober 2022

Revision Date:

15 Oktober 2022

Date Received:

25 Oktober 2022

ABSTRAK

Klinik merupakan fasilitas tempat orang berobat dan memperoleh advis medis serta tempat mahasiswa kedokteran melakukan pengamatan terhadap kasus penyakit yang diderita para pasien. Saat ini, hadirnya virus Corona (COVID-19) membuat banyak klinik menampung pasien yang terpapar virus tersebut. Dari kasus tersebut, rekomendasi penentuan klinik sangat diperlukan karena kondisi yang sangat darurat dan kasus positif yang bertambah setiap harinya. Pada penelitian ini, ditambahkan metode information retrieval, yaitu metode TF-IDF dan BM25 untuk menentukan rekomendasi klinik di Kabupaten Jember berdasarkan kata pencarian dari penggunaannya dan diurutkan berdasarkan kemiripan (similarity) dari yang terbesar hingga yang terkecil. Sementara metode Haversine Formula digunakan untuk memilih klinik dengan jarak yang ditentukan oleh pengguna sebelumnya. Penentuan rekomendasi klinik yang menggunakan metode gabungan information retrieval (similarity) + haversine dilakukan dengan formulasi rata-rata peringkat antara metode haversine dengan metode gabungan, dan formulasi normalisasi nilai similarity maupun nilai haversine. Hasilnya, ada 7 klinik yang menempati peringkat terbaik untuk metode gabungan dengan formulasi rata-rata peringkat, dan ada 47 klinik yang menempati peringkat terbaik untuk metode gabungan dengan formulasi normalisasi.

Keywords: Haversine Formula, Information Retrieval, Okapi BM25, Sistem Rekomendasi, TF-IDF



PENDAHULUAN

Klinik merupakan fasilitas tempat orang berobat dan memperoleh advis medis serta tempat mahasiswa kedokteran melakukan pengamatan terhadap kasus penyakit yang diderita para pasien. Saat ini, hadirnya virus Corona (COVID-19) membuat banyak klinik menampung pasien yang terpapar virus tersebut. Virus yang pertama kali teridentifikasi di kota Wuhan, Cina pada Desember 2019 lalu, telah menularkan ke banyak orang di seluruh dunia, termasuk Indonesia.

Dari kasus tersebut, rekomendasi penentuan klinik sangat diperlukan karena kondisi yang sangat darurat dan kasus positif yang bertambah setiap harinya. Penentuan rekomendasi dapat dilakukan dengan beberapa pendekatan, salah satunya yaitu content-based system. Pendekatan tersebut menerapkan kemiripan atribut dari suatu item terhadap item lain yang direkomendasikan [1]. Atribut dari suatu item dapat berupa deskripsi teks, dengan menerapkan metode information retrieval untuk menilai kemiripannya.

Metode information retrieval yang digunakan untuk menilai kemiripan atribut suatu item yaitu Term Frequency-Inverse Document Frequency (TF-IDF) dan BM25. Metode TF-IDF dilakukan dengan memberi bobot suatu kata untuk menentukan seberapa penting suatu kata terhadap dokumen yang terdapat pada corpus [2]. Sementara metode BM25 dilakukan dengan mengurutkan hasil kecocokan terhadap suatu dokumen berdasarkan kata kunci (query) yang dicari [3].

Oleh karena itu, dalam penelitian ini akan ditambahkan metode information retrieval, yaitu metode TF-IDF dan BM25 untuk menentukan rekomendasi klinik di Kabupaten Jember berdasarkan kata pencarian dari penggunaannya dan diurutkan berdasarkan kemiripan (similarity) dari yang terbesar hingga yang terkecil. Sementara metode Haversine Formula digunakan untuk memilih klinik dengan jarak yang ditentukan oleh pengguna sebelumnya.

Tinjauan Pustaka

Pada tahun 2017, Monica Wideasri, Ellysa Tjandra, dan Lisa Maria Chandra melakukan penelitian menggunakan metode BM25 meningkatkan kinerja pencarian dokumen tugas akhir, yaitu dengan menggunakan 60 query untuk perhitungan nilai precision dan recall untuk setiap query tersebut. 30 dari query merupakan beberapa kata dasar dan variasi kata yang menggunakan kata dasar, yang meliputi: ajar, program, buat, putus, hubung, angkat, kenal, dan dukung. Sedangkan 30 query yang lainnya merupakan query berupa kata yang tidak mempunyai variasi kata dasar yang sama atau query yang terdiri dari 2 kata. Penelitian tersebut menghasilkan nilai recall yang sangat baik yaitu 1 dan nilai precision yang baik yaitu 0,986. Query yang memiliki nilai precision kurang dari rata-rata, dokumen yang relevan tetap berada di urutan atas dan dokumen yang tidak relevan di urutan bawah [4].

Raditya Adi Baskara pada 2017 meneliti metode TF-IDF untuk memberikan rekomendasi tempat makan berdasarkan kata pencarian (keyword) dari pengguna dan diurutkan berdasarkan hasil yang memiliki keterkaitan tertinggi ke yang lebih rendah. Metode tersebut menghasilkan nilai precision sebesar 0,65, yang artinya kemampuan sistem untuk menampilkan restoran yang relevan pada hasil pencarian adalah cukup baik walaupun masih banyak restoran yang kurang relevan yang ditampilkan. Sementara untuk nilai recall yang diperoleh adalah sebesar 0,97, yang berarti sistem dapat menampilkan seluruh hasil relevan dari database [5].

Pada tahun 2015, Ryan Herawan Dwi Putra, Herry Sujiani, dan Novi Safriadi melakukan penelitian menggunakan Haversine Formula untuk pengukuran luas tanah. Pengujian akurasi pada penelitian tersebut dilakukan sebanyak 3 macam, dengan membandingkan metode Haversine Formula dengan GPS dan kenyataan. Pengujian pertama yaitu akurasi panjang, yang hasilnya selisih panjang antara 2 titik dengan menggunakan Haversine Formula dan GPS sebesar 3,33% dan 7,33 jika dibandingkan dengan kenyataan. Pengujian kedua yaitu akurasi luas, yang hasilnya selisih luas dengan menggunakan Haversine Formula dan GPS sebesar 3,923% dan 7,846% jika dibandingkan dengan kenyataan. Pengujian ketiga yaitu pengujian selisih letak pada GPS, yang menggunakan 15 titik untuk dihitung selisih letaknya pada GPS dengan kenyataan. Selisih letak pada GPS dengan kenyataan yaitu 2 meter [6].

Fokus utama dari penelitian ini yaitu penambahan metode information retrieval, yaitu metode TF-IDF dan BM25 untuk menentukan rekomendasi klinik di Kabupaten Jember berdasarkan kata pencarian dari penggunaannya dan diurutkan berdasarkan kemiripan (similarity) dari yang terbesar

hingga yang terkecil. Sementara metode Haversine Formula digunakan untuk memilih klinik dengan jarak yang ditentukan oleh pengguna sebelumnya.

A. Haversine Formula

Haversine Formula yaitu metode yang digunakan untuk menghitung jarak antar dua titik dengan memperhitungkan bahwa bumi bukanlah merupakan bidang yang datar, melainkan bidang yang memiliki derajat kelengkungan. Metode ini menghitung jarak antara dua titik dengan berdasarkan panjang garis lurus antara dua titik yang berada pada garis lintang dan garis bujur [7]. Rumus perhitungan dari metode Haversine Formula ditunjukkan pada persamaan (1)

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (1)$$

Diketahui d adalah jarak dua titik dan r adalah jari-jari bumi (6,371 km) [8]. φ_1 dan φ_2 adalah titik lintang dan λ_1 dan λ_2 adalah titik bujur.

B. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF yaitu metode yang digunakan untuk memberi bobot pada sebuah term yang dilakukan dengan cara mengukur tingkat kepentingan dari suatu term untuk sebuah dokumen teks dalam kumpulan beberapa dokumen teks [9]. Term Frequency (TF) digunakan untuk mengukur seberapa sering suatu kata muncul dalam suatu dokumen. Suatu kata dapat muncul lebih dari satu kali apabila dokumen yang digunakan lebih panjang karena panjang (jumlah kata) dalam suatu dokumen berbeda-beda [2]. Rumus dari perhitungan TF ditunjukkan pada persamaan (2).

$$tf(t, d) = \begin{cases} 1 + \log(f_{t,d}) & \text{if } f_{t,d} > 0 \\ 0 & \text{if } f_{t,d} = 0 \end{cases} \quad (2)$$

Diketahui $tf(t, d)$ merupakan bobot term frequency dan $f_{t,d}$ adalah banyaknya kata t pada dokumen d .

Inverse Documents Frequency (IDF) digunakan untuk mengukur tingkat kepentingan dari kata tersebut. Pada saat menghitung TF, semua kata dianggap penting, namun kata-kata keterangan dan kata yang memiliki frekuensi kemunculan tinggi atau terdapat di hampir seluruh dokumen perlu diseimbangkan dengan yang jarang (frekuensi kemunculan rendah). Rumus dari perhitungan IDF ditunjukkan pada persamaan (3).

$$idf(t, D) = \log \frac{N}{df_t} \quad (3)$$

Diketahui N adalah total banyaknya dokumen dalam sekumpulan dokumen dan df_t adalah banyaknya dokumen yang terdapat kata t , apabila kata tersebut bernilai tidak ada pada seluruh dokumen (0) maka nilainya adalah 1.

Berdasarkan perhitungan TF dan IDF yang telah dijabarkan pada persamaan (2) sampai (3), maka nilai TF-IDF dihitung dengan mengalikan keduanya. Rumus untuk menghitung TF-IDF ditunjukkan pada persamaan (4).

$$tf \cdot idf(t, d, D) = (1 + \log(f_{t,d})) \left(\log \frac{N}{df_t} \right) \quad (4)$$

Nilai dari TF-IDF dapat berpengaruh sesuai dengan frekuensi kemunculan term pada suatu dokumen dan diimbangi dengan jumlah dokumen dalam korpus yang berisi term tersebut. Nilai dapat meningkat secara proporsional sehingga membantu menyesuaikan fakta bahwa beberapa term dapat sering muncul secara umum [10].

C. Cosine Similarity

Cosine similarity yaitu metode yang digunakan untuk mengukur kemiripan di antara dua dokumen dengan cara menjadikan atribut dari dokumen ke dalam vektor, kemudian dihitung sudut kosinus di antara kedua vektor tersebut [11]. Penghitungan dari metode cosine similarity didefinisikan pada persamaan (5).

$$\text{CosineSimilarity}(d_j, q) = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij} \cdot \sum_{i=1}^t w_{iq}}} \quad (5)$$

Diketahui (d_j, q) merupakan nilai similarity dari dokumen j dengan query q , sementara w_{ij} adalah nilai bobot dari term i ke dokumen j dan w_{iq} nilai bobot dari term i ke query q .

D. Okapi BM25

Okapi BM25 yaitu metode fungsi pemeringkatan yang dibuat oleh Stephen Robertson dan Karen Sparck Jones, yang digunakan untuk mengurutkan hasil kemiripan (similarity) terhadap seluruh dokumen terhadap kata kunci (query). Metode Okapi BM25 efektif dan memiliki ketepatan untuk mengurutkan dokumen berdasarkan query, sehingga metode tersebut merupakan metode terbaik dalam kelas best match [12]. Skor dari Okapi BM25 dipengaruhi oleh tiga faktor, yaitu TF (term frequency), IDF (inverse document frequency) dan panjang dokumen berdasarkan jumlah kata. Dokumen yang singkat, namun di dalamnya terdapat query yang dicari pengguna jauh lebih baik daripada dokumen yang memiliki jumlah kata lebih banyak namun tidak terdapat query [13]. Perhitungan skor dari Okapi BM25 ditunjukkan pada persamaan (6).

$$\text{BM25} = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left((1 - b) + b \cdot \frac{|D|}{\text{avgdl}} \right)} \quad (6)$$

Diketahui $f(q_i, D)$ adalah jumlah kemunculan kata q_i pada dokumen D , $|D|$ adalah jumlah kata pada dokumen D dan avgdl is rata-rata panjang dokumen dari sekumpulan dokumen, sementara k_1 adalah 1,2 dan b adalah 0,75.

Variabel k_1 adalah parameter tuning positif yang menyesuaikan scaling term frequency pada dokumen. $k_1 = 0$ artinya tidak ada term frequency. Sementara variabel b adalah parameter tuning lain yang menentukan scaling dari panjang dokumen. $b = 0$ artinya tidak ada normalisasi untuk panjang dokumen [14]. Fungsi $\text{IDF}(q_i)$ sebagaimana pada persamaan (6), dijabarkan pada persamaan (7).

$$\text{IDF}(q_i) = \log \left(\frac{N - n(q_i) + 0,5}{n(q_i) + 0,5} \right) \quad (7)$$

Diketahui $n(q_i)$ adalah jumlah dokumen yang mengandung term q_i dan N jumlah dokumen dalam koleksi.

E. Google Maps API

Google Maps Application Program Interface (API) adalah sebuah layanan pemetaan yang disediakan oleh Google. API merupakan suatu set aturan dan spesifikasi tertentu yang menyatakan bagaimana program mengakses dan memanfaatkan sumber daya yang disediakan oleh program lainnya yang menggunakan API. API dapat dinyatakan sebagai penghubung antara satu program dengan program lainnya [15].

Google Maps (tanpa API) mulai diperkenalkan pada bulan Februari 2005 yang merupakan revolusi bagaimana peta di dalam halaman web dengan membiarkan pengguna untuk menarik peta sehingga dapat menavigasinya. Pada saat itu, solusi peta ini masih baru dan membutuhkan server khusus.

Beberapa saat kemudian, ada yang berhasil membobol Google Maps untuk digunakan pada webnya sendiri. Oleh karena itu, Google Maps mengambil keputusan bahwa Google Maps membutuhkan API dan pada bulan Juni 2005, Google Maps API dirilis secara publik [16].

METODE

Penelitian ini menggunakan dua tahapan pengerjaan. Tahapan yang pertama adalah mempersiapkan data peta, dan tahapan yang kedua adalah menggunakan data peta yang telah dipersiapkan untuk menentukan rekomendasi klinik menggunakan metode information retrieval dan Haversine Formula. Diagram alir dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Metodologi Penelitian

Dataset yang digunakan pada penelitian ini yaitu data berupa fasilitas kesehatan (faskes) dari masing-masing kecamatan yang ada di Kabupaten Jember, diantaranya data faskes rumah sakit, faskes rumah sakit TNI/POLRI, dan data faskes puskesmas. Data tersebut diperoleh secara online pada situs Bantuan BPJS, yang berisi kode faskes, nama faskes dan alamat faskes yang ada di Kabupaten Jember [17].

Penelitian ini menggunakan Haversine Formula dengan menentukan titik pertama sebagai titik lokasi pengguna, dan titik kedua sebagai titik lokasi klinik tujuan. Kemudian, dilakukan proses perhitungan metode Haversine Formula. Setelah mendapatkan hasil perhitungan, maka penentuan rekomendasi ditentukan dengan mengurutkan jarak dari titik awal dan tujuan berdasarkan jarak terdekat.

Penelitian ini juga menggunakan metode information retrieval dengan penghitungan nilai TF-IDF dan Okapi BM25 berdasarkan kata kunci yang dimasukkan oleh pengguna. Tahap awal yaitu dengan melakukan data preprocessing, lalu menghitung pembobotan dari TF-IDF maupun Okapi BM25. Setelah masing-masing pembobotan dilakukan, tahapan berikutnya yaitu penghitungan nilai cosine similarity dan mengurutkannya berdasarkan nilai similarity tertinggi untuk menentukan rekomendasi klinik sesuai kata kunci yang dimasukkan pengguna.

Lalu penelitian ini akan membandingkan pengurutan rekomendasi klinik dengan menggunakan Haversine Formula method dan dengan metode yang diusulkan dalam penelitian ini, yaitu gabungan Haversine Formula dan Information Retrieval.

HASIL DAN PEMBAHASAN

Sebelum mengimplementasikan metode, diawali dengan menyiapkan data peta mengenai klinik di Kabupaten Jember. Data peta berisi nama klinik, alamat, kecamatan, fasilitas, and titik koordinat (titik lintang dan titik bujur) yang ditentukan oleh Google Maps API.

Implementasi Metode Information Retrieval

Pada sistem rekomendasi yang dibangun, metode information retrieval menerapkan proses pencarian dengan kata kunci yang ditentukan oleh pengguna. Data yang digunakan sebagai uji coba metode tersebut yaitu data faskes dari masing-masing kecamatan di Kabupaten Jember, yang memuat nama, alamat, kecamatan, dan fasilitas yang tersedia dari tiap faskes. Kata kunci yang dimasukkan yaitu "sakit jantung".

Setelah dilakukan tokenisasi, langkah selanjutnya yaitu menghitung pembobotan dari metode TF-IDF dan Okapi BM25. Pada umumnya, kedua metode perhitungan tersebut memiliki kemiripan. Pada metode Okapi BM25, yang membedakan antara metode TF-IDF yaitu penghitungan nilai IDF dan penghitungan bobot dari metode tersebut.

Setelah menghitung pembobotan, maka tahapan selanjutnya yaitu menghitung kemiripan antara kata kunci terhadap masing-masing data faskes, dengan menggunakan Cosine Similarity. Setelah itu, data diurutkan berdasarkan nilai cosine similarity dan rating masing-masing data faskes yang diperoleh dari ulasan pada Google Maps, sehingga menghasilkan rekomendasi klinik berdasarkan kata kunci yang dimasukkan. Hasil pengurutan data dengan metode TF-IDF ditunjukkan pada Tabel 1, sementara pengurutan dengan metode Okapi BM25 ditunjukkan pada Tabel 2.

Tabel 1. Urutan Data Faskes Berdasarkan Nilai Cosine Similarity (TF-IDF) dan Rating

Nama Klinik	Rating	Similarity	Ranking
Rumah Sakit Daerah Kalisat	3.9	0.256261974	1
Rumah Sakit Citra Husada	4.3	0.236453338	2
Rumah Sakit Paru	4.4	0.199068052	3
Rumah Sakit Daerah Dr Soebandi Jember	3.9	0.191799148	4
Rumah Sakit Bina Sehat	3.4	0.177195029	5
Rumah Sakit Umum Kaliwates	3.8	0.125639831	6
Rumah Sakit DKT Baladhika Husada Jember	3.9	0.121426865	7
Rumah Sakit Perkebunan PTPN X	3.5	0.09781322	8
Rumah Sakit Ibu dan Anak Srikandi IBI	4.3	0.092097218	9
Rumah Sakit Daerah Balung	3	0.083849836	10

Sumber: Data diolah (2022)

Tabel 2. Urutan Data Faskes Berdasarkan Nilai Cosine Similarity (Okapi BM25) dan Rating

Nama Klinik	Rating	Similarity	Ranking
Rumah Sakit Daerah Kalisat	3.9	0.193496347	1
Rumah Sakit Citra Husada	4.3	0.170241022	2
Rumah Sakit Paru	4.4	0.16780792	3
Rumah Sakit Daerah Dr Soebandi Jember	3.9	0.149362985	4
Rumah Sakit Bina Sehat	3.4	0.136415194	5
Rumah Sakit Umum Kaliwates	3.8	0.081900865	6
Rumah Sakit DKT Baladhika Husada Jember	3.9	0.07897537	7
Rumah Sakit Perkebunan PTPN X	3.5	0.066202965	8
Rumah Sakit Ibu dan Anak Srikandi IBI	4.3	0.065606147	9
Rumah Sakit Daerah Balung	3	0.060313611	10

Sumber: Data diolah (2022)

Implementasi Metode Haversine Formula

Penerapan metode Haversine Formula dilakukan dengan menentukan titik latitude dan longitude pertama sebagai titik lokasi awal, dan titik latitude dan longitude kedua sebagai titik lokasi klinik tujuan. Sebagai contoh dilakukan penghitungan jarak dari alun-alun kota Jember untuk titik awalnya, dengan titik latitude -8.1686391000 dan titik longitude 113.7020076000 terhadap seluruh klinik seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Titik Latitude dan Longitude Klinik

No	Nama Klinik	Latitude	Longitude
1	Rumah Sakit Citra Husada	-8.1664084075	113.6805809163
2	Rumah Sakit Perkebunan PTPN X	-8.1680342465	113.7050909817
3	Rumah Sakit Ibu dan Anak Srikandi IBI	-8.1858022272	113.6903573471
4	Rumah Sakit Umum Kaliwates	-8.1814330679	113.6751362840
5	Rumah Sakit Daerah Dr Soebandi Jember	-8.1509412865	113.7155041236
6	Rumah Sakit Daerah Balung	-8.2709133871	113.5399165967
7	Rumah Sakit Daerah Kalisat	-8.1337614478	113.8214452256
8	Rumah Sakit Paru	-8.1601839352	113.7063115644
9	Rumah Sakit Bina Sehat	-8.1803965537	113.6849253302
10	Rumah Sakit DKT Baladhika Husada Jember	-8.1638959876	113.7061781786
...
58	Puskesmas Pakusari	-8.1772457268	113.7783941151
59	Puskesmas Sukorambi	-8.1589539941	113.6661017256

Sumber: Data diolah (2022)

Dari data di atas, jarak diukur dengan menggunakan metode Haversine Formula, kemudian diurutkan dari jarak terpendek ke jarak terpanjang, untuk menentukan rekomendasi klinik berdasarkan jarak dua titik. Hasil urutan jarak berdasarkan metode Haversine Formula ditunjukkan pada Tabel 4.

Tabel 4. Hasil Urutan Jarak Berdasarkan Haversine Formula

Nama Klinik	Distance (km)	Ranking
Rumah Sakit Perkebunan PTPN X	0.3459782	1
Rumah Sakit DKT Baladhika Husada Jember	0.699202	2
Rumah Sakit Paru	1.0527777	3
Puskesmas Sumbersari	1.2328202	4
Puskesmas Patrang	1.4866142	5
Puskesmas Jember Kidul	1.7915536	6
Rumah Sakit Bina Sehat	2.2900272	7
Rumah Sakit Ibu dan Anak Srikandi IBI	2.2992234	8
Rumah Sakit Citra Husada	2.3713802	9
Rumah Sakit Daerah Dr Soebandi Jember	2.4656686	10

Sumber: Data diolah (2022)

Implementasi Metode Gabungan Similarity dan Haversine

Penentuan peringkat rekomendasi untuk metode gabungan similarity dan haversine dilakukan dengan formulasi skor ranking berupa formulasi rata-rata peringkat metode similarity dengan metode haversine, dan formulasi normalisasi nilai similarity maupun nilai haversine (jarak). Formulasi skor dengan rata-rata peringkat dilakukan dengan mencari rata-rata peringkat kedua metode (haversine saja dan gabungan), lalu diurutkan berdasarkan hasil rata-ratanya. Hasil urutan rekomendasi metode gabungan dengan formulasi rata-rata peringkat ditunjukkan pada tabel 5.

Tabel 5. Hasil Urutan Rekomendasi Metode Gabungan Dengan Formulasi Rata-Rata Peringkat

Nama Klinik	Ranking Similarity	Ranking Haversine	Rata-rata Ranking	Ranking Final
Rumah Sakit Paru	3	3	3	1
Rumah Sakit Perkebunan PTPN X	8	1	4,5	2
Rumah Sakit DKT Baladhika Husada Jember	7	2	4,5	3
Rumah Sakit Citra Husada	2	9	5,5	4
Rumah Sakit Bina Sehat	5	7	6	5
Rumah Sakit Daerah Dr Soebandi Jember	4	10	7	6
Puskesmas Sumbersari	11	4	7,5	7
Puskesmas Patrang	12	5	8,5	8
Rumah Sakit Ibu dan Anak Srikandi IBI	9	8	8,5	9
Rumah Sakit Umum Kaliwates	6	11	8,5	10
...
Puskesmas Cakru	58	58	58	58
Puskesmas Jombang	59	59	59	59

Sumber: Data diolah (2022)

Formulasi skor dengan normalisasi dilakukan dengan menormalisasi nilai similarity dan nilai haversine (jarak) dengan rentang nilai antara 0 sampai 100. Similarity yang dinormalisasi bernilai 100 untuk similarity tertinggi, sementara jarak yang dinormalisasi bernilai 100 untuk jarak terpendek. Kemudian, skor pemeringkatan ditentukan dengan mencari selisih nilai normalisasi antara jarak dengan similarity, dan skor pemeringkatan berupa nilai mutlak, Urutan rekomendasi dengan formulasi ini ditentukan dari hasil skor tertinggi ke terendah. Hasil urutan rekomendasi ditunjukkan pada Tabel 6.

Tabel 6. Hasil Urutan Rekomendasi Metode Gabungan Dengan Formulasi Normalisasi

Nama Klinik	Normalisasi		Skor Ranking	Ranking Final
	Similarity	Jarak (KM)		
Puskesmas Sumbersari	0	97,91348217	97,91348217	1
Puskesmas Patrang	0	97,316368	97,316368	2
Puskesmas Jember Kidul	0	96,59892179	96,59892179	3
Puskesmas Sukorambi	0	91,17667558	91,17667558	4
Puskesmas Kaliwates	0	91,13514508	91,13514508	5
Puskesmas Gladak Pakem	0	89,83944032	89,83944032	6
Puskesmas Mangli	0	86,17647781	86,17647781	7
Puskesmas Panti	0	81,14898331	81,14898331	8
Puskesmas Pakusari	0	80,90545401	80,90545401	9
Puskesmas Ajung	0	80,1800122	80,1800122	10
...
Puskesmas Cakru	0	1,208529025	1,208529025	58
Puskesmas Jombang	0	0	0	59

Sumber: Data diolah (2022)

Perbandingan Urutan Rekomendasi

Pada subbab ini, dibahas mengenai perbandingan peringkat rekomendasi klinik antara metode haversine saja dengan metode gabungan antara similarity dan haversine, yang telah ditentukan peringkat akhirnya pada subbab sebelumnya. Perbandingan peringkat rekomendasi beserta peringkat

terbaik ditunjukkan pada tabel 7, sementara jumlah klinik yang menempati peringkat terbaik dari masing-masing metode ditunjukkan pada tabel 8.

Tabel 7. Perbandingan Peringkat Rekomendasi

Nama Klinik	Ranking Haversine	Ranking Metode Gabungan		Ranking Terbaik	
		Rata-rata	Normalisasi	Rata-rata	Normalisasi
Rumah Sakit Perkebunan PTPN X	1	2	20	Haversine	Haversine
Rumah Sakit DKT Baladhika Husada Jember	2	3	30	Haversine	Haversine
Rumah Sakit Paru	3	1	50	Similarity + Haversine	Haversine
Puskesmas Sumbersari	4	7	1	Haversine	Similarity + Haversine
Puskesmas Patrang	5	8	2	Haversine	Similarity + Haversine
Puskesmas Jember Kidul	6	11	3	Haversine	Similarity + Haversine
Rumah Sakit Bina Sehat	7	5	47	Similarity + Haversine	Haversine
Rumah Sakit Ibu dan Anak Srikandi IBI	8	9	21	Haversine	Haversine
Rumah Sakit Citra Husada	9	4	57	Similarity + Haversine	Haversine
Rumah Sakit Daerah Dr Soebandi Jember	10	6	51	Similarity + Haversine	Haversine
...
Puskesmas Cakru	58	58	58	Sama	Sama
Puskesmas Jombang	59	59	59	Sama	Sama

Sumber: Data diolah (2022)

Tabel 8. Jumlah Klinik Penempatan Peringkat Terbaik

Metode Pemingkatan Terbaik	Jumlah Klinik		Persentase	
	Rata-rata	Normalisasi	Rata-rata	Normalisasi
Haversine	29	10	49,15%	16,95%
Similarity + Haversine	7	47	11,87%	79,66%
Sama	23	2	38,98%	3,39%

Sumber: Data diolah (2022)

Dari tabel 8, untuk penggunaan metode gabungan similarity + haversine dengan formulasi rata-rata peringkat, hanya ada 7 klinik yang menempati peringkat terbaik untuk metode gabungan, 29 klinik untuk metode haversine, dan 23 klinik berada pada peringkat yang sama. Sementara penggunaan metode gabungan dengan formulasi normalisasi, terdapat 47 klinik yang menempati peringkat terbaik untuk metode gabungan, 10 klinik untuk metode haversine, dan 2 klinik berada pada peringkat yang sama. Maka, pada penelitian ini, formulasi terbaik untuk penggabungan metode similarity + haversine yaitu dengan cara normalisasi.

Evaluasi Metode Information Retrieval

Evaluasi yang dilakukan dalam penelitian ini mengukur precision, recall, dan F1-score. Pengujian ini dilakukan dengan memasukkan kata kunci sebagai pencarian pada data klinik. Kata kunci yang dimasukkan pada percobaan ini antara lain “sakit jantung”, “paru paru kronis”, “gangguan telinga”,

“gigi berlubang”, “kebidanan”, “sakit kulit”, “sakit panu”, “panu”, “gangguan mata”, dan “sakit perut”. Dari sepuluh kata kunci tersebut, Dari kesepuluh kata kunci tersebut, hasil dari percobaan yang dilakukan ditunjukkan pada Tabel 9.

Tabel 9. Hasil Pengujian Precision, Recall, Dan F1-Score

Keyword	Precision	Recall	F1-Score
1	0.5	1	0.67
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	0,2	1	0,33
7	0	0	0
8	0	0	0
9	1	1	1
10	0	0	0
Average	0,57	0,7	0,6

Sumber: Data diolah (2022)

Pada evaluasi performa dari metode information retrieval, didapatkan nilai precision yang diperoleh yaitu sebesar 0,57, nilai recall sebesar 0,7, dan nilai F1-score sebesar 0,6. Fasilitas pelayanan kesehatan tiap klinik yang kurang spesifik dapat menjadi penyebab nilai evaluasi tersebut cukup rendah. Pada data klinik yang diperoleh, fasilitas pelayanan kesehatan hanya secara umum seperti poli umum, poli kulit, poli gigi, dan sebagainya, sehingga untuk kata kunci penyakit secara khusus, seperti “sakit panu”, “panu”, dan “sakit perut” dianggap tidak relevan.

Evaluasi Metode Haversine Formula

Pengujian ini dilakukan dengan menggunakan titik letak klinik dengan posisi awal yaitu alun-alun kota Jember, yang terletak pada titik (-8.1686391000, 113.7020076000). Pengujian dilakukan dengan menghitung selisih dari jarak hasil pengukuran pada Haversine Formula dengan pengukuran garis lurus maupun jarak secara nyata pada Google Maps API. Hasil pengujian pengukuran Haversine Formula dengan garis lurus pada Google Maps API ditunjukkan pada Tabel 10.

Tabel 10. Pengujian Pengukuran Haversine Formula Dengan Pengukuran Garis Lurus Pada Google Maps API

No	Nama Klinik	Haversine (km)	Line Measure (km)	Difference (Km)	Difference (%)
1	Rumah Sakit Citra Husada	2.371	2.380	0.009	0.362
2	Rumah Sakit Perkebunan PTPN X	0.346	0.325	0.021	6.455
3	Rumah Sakit Ibu dan Anak Srikandi IBI	2.299	2.300	0.001	0.034
4	Rumah Sakit Umum Kaliwates	3.282	3.310	0.028	0.847
5	Rumah Sakit Daerah Dr Soebandi Jember	2.466	2.490	0.024	0.977
6	Rumah Sakit Daerah Balung	21.155	21.150	0.005	0.025
7	Rumah Sakit Daerah Kalisat	13.707	13.730	0.023	0.169
8	Rumah Sakit Paru	1.053	1.080	0.027	2.521
9	Rumah Sakit Bina Sehat	2.290	2.300	0.010	0.434
10	Rumah Sakit DKT Baladhika Husada Jember	0.699	0.721	0.022	3.023

...
58	Puskesmas Pakusari	8.462	8.440	0.022	0.259
59	Puskesmas Sukorambi	4.096	4.110	0.014	0.336
Average				0.645	0.639

Sumber: Data diolah (2022)

Pada pengujian Haversine Formula dengan pengukuran garis lurus pada Google Maps API, terdapat selisih perhitungan yaitu sebesar 0,645 km atau 645 meter, sementara persentasenya yaitu 0,639% atau nilai akurasinya 99,361%. Selisih tersebut terjadi karena adanya pembulatan pada titik latitude maupun longitude pada Google Maps API. Hasil pengujian pengukuran Haversine Formula dengan jarak nyata pada Google Maps API ditunjukkan pada Tabel 11.

Tabel 11. Pengujian Pengukuran Haversine Formula Dengan Pengukuran Jarak Nyata Pada Google Maps API

No	Nama Klinik	Haversine (km)	Real Distance (km)	Difference (Km)	Difference (%)
1	Rumah Sakit Citra Husada	2.371	5.300	2.929	55.257
2	Rumah Sakit Perkebunan PTPN X	0.346	0.450	0.104	23.116
3	Rumah Sakit Ibu dan Anak Srikandi IBI	2.299	3.700	1.401	37.859
4	Rumah Sakit Umum Kaliwates	3.282	4.900	1.618	33.021
5	Rumah Sakit Daerah Dr Soebandi Jember	2.466	3.100	0.634	20.462
6	Rumah Sakit Daerah Balung	21.155	23.400	2.245	9.593
7	Rumah Sakit Daerah Kalisat	13.707	17.800	4.093	22.996
8	Rumah Sakit Paru	1.053	1.600	0.547	34.201
9	Rumah Sakit Bina Sehat	2.290	4.000	1.710	42.749
10	Rumah Sakit DKT Baladhika Husada Jember	0.699	1.100	0.401	36.436
...
58	Puskesmas Pakusari	8.462	10.400	1.938	18.636
59	Puskesmas Sukorambi	4.096	7.700	3.604	46.803
Average				3.994	22.411

Sumber: Data diolah (2022)

Pada pengujian Haversine Formula dengan pengukuran jarak nyata pada Google Maps API, terdapat selisih perhitungan yaitu sebesar 3,994 km atau 3.994 meter, sementara persentasenya yaitu 22,411% atau nilai akurasinya 77,589%. Selisih tersebut cukup besar karena pengukuran jarak nyata pada Google Maps juga diperhitungkan dari sisi belokan, dan rute yang ditempuh menyesuaikan dengan jalur yang diperbolehkan untuk melintas, sesuai dengan peraturan lalu lintas yang berlaku. Sementara pada pengukuran Haversine Formula, hanya ditentukan dari jarak garis lurus antara dua titik. Selain itu, selisih yang cukup besar juga disebabkan oleh pengukuran jarak pada metode Haversine Formula tidak memperhatikan topografi dari permukaan tanah, seperti pegunungan dan lembah.

KESIMPULAN

Penentuan rekomendasi klinik yang menerapkan metode gabungan information retrieval (similarity) dan Haversine Formula dilakukan dengan memasukkan kata kunci dan menentukan titik lokasi awal

untuk menghitung jarak antara titik awal dengan titik tujuan masing-masing lokasi klinik. Pada penelitian ini, kata kunci yang digunakan yaitu “sakit jantung” dan titik awal untuk pengukuran jarak yaitu alun-alun kota Jember. Penentuan peringkat dari metode gabungan dilakukan dengan formulasi rata-rata peringkat metode similarity dengan metode haversine, dan formulasi normalisasi nilai similarity maupun nilai haversine. Untuk penggunaan metode gabungan similarity + haversine dengan formulasi rata-rata peringkat, hanya ada 7 klinik yang menempati peringkat terbaik untuk metode gabungan, 29 klinik untuk metode haversine, dan 23 klinik berada pada peringkat yang sama. Sementara penggunaan metode gabungan dengan formulasi normalisasi, terdapat 47 klinik yang menempati peringkat terbaik untuk metode gabungan, 10 klinik untuk metode haversine, dan 2 klinik berada pada peringkat yang sama. Formulasi terbaik untuk penggabungan metode similarity + haversine yaitu dengan formulasi normalisasi.

Pada evaluasi performa dari metode information retrieval, didapatkan nilai precision yang diperoleh yaitu sebesar 0,57, nilai recall sebesar 0,7, dan nilai F1-score sebesar 0,6. Fasilitas pelayanan kesehatan tiap klinik yang kurang spesifik dapat menjadi penyebab nilai evaluasi tersebut cukup rendah. Pada data klinik yang diperoleh, fasilitas pelayanan kesehatan hanya secara umum seperti poli umum, poli kulit, poli gigi, dan sebagainya, sehingga untuk kata kunci penyakit secara khusus, seperti “sakit panu”, “panu”, dan “sakit perut” dianggap tidak relevan.

Pada evaluasi performa dari metode Haversine Formula, pengujian dilakukan dengan dua cara. Yang pertama, perbandingan Haversine terhadap pengukuran garis lurus pada Google Maps API, terdapat selisih perhitungan yaitu sebesar 0,645 km atau 645 meter, sementara persentasenya yaitu 0,639% atau nilai akurasi 99,361%. Selisih tersebut terjadi karena adanya pembulatan pada titik latitude maupun longitude pada Google Maps API. Yang kedua, perbandingan Haversine terhadap pengukuran jarak nyata pada Google Maps API, terdapat selisih perhitungan yaitu sebesar 3,994 km atau 3.994 meter, sementara persentasenya yaitu 22,411% atau nilai akurasi 77,589%. Selisih tersebut cukup besar karena pengukuran jarak nyata pada Google Maps juga diperhitungkan dari sisi belokan, dan rute yang ditempuh menyesuaikan dengan jalur yang diperbolehkan untuk melintas, sesuai dengan peraturan lalu lintas yang berlaku. Sementara pada pengukuran Haversine Formula, hanya ditentukan dari jarak garis lurus antara dua titik. Selain itu, selisih yang cukup besar juga disebabkan oleh pengukuran jarak pada metode Haversine Formula tidak memperhatikan topografi dari permukaan tanah, seperti pegunungan dan lembah.

Saran yang dapat dilakukan untuk perbaikan maupun pengembangan pada penelitian selanjutnya yaitu penambahan informasi dari data fasilitas kesehatan untuk meningkatkan nilai performa pada pencarian yang lebih relevan. Saran berikutnya yaitu menambahkan fasilitas pelayanan kesehatan dari masing-masing klinik menjadi lebih spesifik agar dapat meningkatkan nilai relevansi sesuai pencarian. Saran lainnya yaitu menentukan klasifikasi penyakit secara khusus sebagai kata kunci ke fasilitas pelayanan kesehatan secara umum, seperti panu termasuk ke dalam poli umum atau poli kulit

REFERENCES

- [1] F. Ricci, L. Rokach, B. Saphira dan P. B. Kantor, *Recommender System Handbook*, Springer, 2011.
- [2] C. D. Manning, P. Raghavan dan H. Schütze, *Scoring, Term Weighting, and The Vector Space Model*, Cambridge University Press, 2008.
- [3] K. Frinta, Indriati dan P. P. Adikara, “Pencarian Berita Berbahasa Indonesia Menggunakan Metode BM25,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2589-2595, 2019.
- [4] M. Widiastri, E. Tjandra dan L. M. Chandra, “Peningkatan Kinerja Pencarian Dokumen Tugas Akhir menggunakan Porter Stemmer Bahasa Indonesia dan Fungsi Peringkat Okapi BM25,” *TEKNIKA*, vol. 6, no. 1, pp. 54-60, 2017.

- [5] R. A. Baskara, Rancang Bangun Sistem Rekomendasi Restoran Berbasis Web menggunakan TF-IDF dan Haversine Formula, Tangerang: Program Studi Teknik Informatika Universitas Multimedia Nusantara, 2017.
- [6] R. H. D. Putra, H. Sujiani dan N. Safriadi, "Penerapan Metode Haversine Formula Pada Sistem Informasi Geografis Pengukuran Luas Tanah," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 1, no. 1, pp. 1-6, 2015.
- [7] G. Brummelen, *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*, Princeton University Press, 2013.
- [8] D. R. Lide, *Handbook of Chemistry and Physics*, CRC Press, 2000.
- [9] J. Chen, C. Chen dan Y. Liang, "Optimized TF-IDF Algorithm with the Adaptive Weight of Position of Word," *Advances in Intelligent Systems Research*, vol. 133, pp. 114-117, 2016.
- [10] M. Fikri, *Pembobotan Term Berbasis Fitur Named Entity dan Posisi Term untuk Klasifikasi Multi-Label Artikel Berita*, Surabaya: Departemen Teknik Informatika Institut Teknologi Sepuluh Nopember, 2019.
- [11] W. H. Goma, "A Survey of Text Similarity Approaches," vol. 68, no. 13, pp. 13-18, 2013.
- [12] J. Pardede, M. G. Husada dan R. Riansyah, "Implementasi Dan Perbandingan Metode Okapi BM25 Dan PLSA Pada Aplikasi Information Retrieval," pp. 1-10, 2018.
- [13] S. Russell dan P. Norvig, *Artificial Intelligence A Modern Approach*, Prentice Hall, 2013.
- [14] C. D. Manning, P. Raghavan dan H. Schütze, *An introduction to information retrieval*, England: Cambridge University Press, 2008.
- [15] Yulianto, Ramadiani dan A. H. Kridalaksana, "Penerapan Formula Haversine pada Sistem Informasi Geografis Pencarian Jarak Terdekat Lokasi Lapangan Futsal," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 13, no. 1, pp. 14-21, 2018.
- [16] G. Svennerberg, *Beginning Google Maps API 3*, New York: Springer Science+Business Media, LLC, 2010.
- [17] "BantuanBPJS.com," [Online]. Available: <https://www.bantuanbpjs.com/alamat-dan-kode-faskes-rumah-sakitpuskesmas-klinik-dokter-dan-apotek-di-kab-jember/>. [Diakses 2 February 2021].