

## **SISTEM TEMU KEMBALI LOKASI WISATA KABUPATEN SUBANG BERBASIS WEB MENGGUNAKAN METODE TERM FREQUENCY - INVERSE DOCUMENT FREQUENCY (TF- IDF)**

Usep Tatang Suryadi<sup>#1</sup>, Zulfiqro<sup>#2</sup>

Program Studi Teknik Informatika, STMIK Subang  
Jl. Marsinu No. 5 - Subang, Tlp. 0206-417853 Fax. 0206-411873  
E-mail : usep\_is\_ugie@yahoo.co.id<sup>#1</sup>, fiqrozul@gmail.com<sup>#2</sup>

### **Abstrak**

Wisata merupakan suatu kegiatan yang telah menjadi salah satu kebutuhan pokok manusia. Objek-objek wisata tersebar di seluruh permukaan bumi dengan berbagai pilihan, mulai dari wisata alam, wahana dan lain sebagainya. Penyebaran objek wisata di Indonesia menjadi salah satu sumber devisa bagi negara umumnya, penda setempat khususnya. Namun semakin luas dan banyak sebaran semakin sulit dalam pemilihan objek-objek wisata tersebut. Di era globalisasi ini, banyak upaya untuk mempermudah pencarian lokasi wisata, dengan peta wisata digital misalnya. Tapi tidak jarang ada beberapa spot wisata yang hanya terekspos melalui artikel. Berupa teks cerita pengalaman perjalanan blogger terkait. Sehingga dalam kesempatan penelitian ini, kami merancang suatu sistem rekomendasi wisata menggunakan sistem temu kembali text mining. Dalam sistem temu kembali sendiri ada beberapa algoritma, dari sekian banyak algoritma kami menggunakan algoritma TF-IDF. Ada beberapa proses pada TF-IDF yaitu Stemming, Stopword Removal, dan Tokenizing. Kemudian menghitung Query term dari dokumen yang ada pada database, lalu melihat jumlah vektor terbesar dari hasil perhitungan. Semakin besar nilai vektor maka dokumen tersebut relevan dengan Query yang diinputkan. Sistem ini telah diuji akurasi dengan perhitungan confusion matrix dengan nilai akurasi sebesar 90%.

**Kata Kunci:** Wisata, Text Mining, Term Frequency Inversed Document Frequency (TF-IDF), Sistem Temu Kembali, Confusion Matrix.

### **Abstract**

The tourism is an activity that has become one of the human primary needs. Tourist spots are spread on various surfaces with various choices, such as nature, playground and many more. The spread of tourism objects in Indonesia becomes one of the members for the country in general, certain local governments. But it is wider and has more distribution in the selection of these tourist spots. In the era of globalization, many efforts to find information, using digital tourism map for example. But sometimes there are some tourist spots are only exposed through articles. In the form of text stories about blogger tour experiences. In this research, we will propose the system using text mining - information retrieval method. In the own information retrieval system there are several algorithms, and we propose the TF-IDF algorithm. There are several processes on TF-IDF such as Stemming, Stopword Removal, and Tokenizing. Then calculate the query term from the documents in the database, then see the largest number of vectors from the calculation results. More higher the vector value, the document is relevant to the input query. This system has calculated its accuracy by confusion matrix method with an accuracy value of 90%.

**Keywords:** Tourist Spots, Text Mining, TF-IDF, Information Retrieval, Confusion Matrix.

## I. PENDAHULUAN

### 1.1 Latar Belakang

Bila dikembangkan dengan sungguh-sungguh, pariwisata di Indonesia memiliki potensi menjadi penyumbang devisa terbesar bagi Indonesia, seperti yang disampaikan oleh Arief Yahya Menteri Pariwisata Indonesia pada Kompas hari ini online 17 Oktober 2017. Peningkatan kebutuhan akan pariwisata dari waktu ke waktu menjadi ukuran perkembangan bidang pariwisata. Tentu peningkatan kebutuhan tersebut perlu diimbangi dengan kemudahan dalam mencari informasi terkait sebaran tempat wisata sesuai dengan kebutuhan berdasarkan jenisnya.

Dalam penggunaan peta wisata sendiri masih banyak tempat-tempat wisata yang masih belum terpetakan. Sebagian tersebar pada blog-blog atau artikel traveler baik yang membahas tempat wisata maupun tentang pengalaman pribadi blogger terkait. Dalam mencari informasi melalui banyak catatan dan artikel membutuhkan waktu yang cukup lama. *Text Mining* adalah bidang baru dan menarik dalam penelitian dunia komputer yang mampu menyelesaikan permasalahan informasi yang berlebih dan merupakan gabungan dari *data mining*, *machine learning*, sistem temu kembali, *knowledge management* dan *natural language processing* (Feldman & Sanger, 2006). Dari beberapa Algoritma Text Mining, TF-IDF adalah satu dari sekian metode text mining yang populer.

### 1.2 Identifikasi Masalah

Dari pemaparan latar belakang, beberapa masalah yang ada diantaranya:

1. Belum adanya sistem rekomendasi pemilihan objek wisata menggunakan *text mining* pada bidang kepariwisataan.
2. Belum diketahui akurasi algoritma tf-idf pada data teks kepariwisataan.

### 1.3 Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan untuk membangun sistem rekomendasi objek wisata berbasis data teks yang bisa membantu wisatawan untuk mencari objek wisata sesuai dengan jenis wisata yang diinginkan, juga mengukur nilai akurasi metode tf-idf. Adapun manfaat dari penelitian ini menambah kepustakaan penerapan bidang komputer pada berbagai aspek kehidupan untuk referensi penelitian selanjutnya. Juga menerapkan metode yang memiliki akurasi yang baik.

### 1.4 Batasan Masalah

Untuk efektivitas waktu, penelitian ini dibatasi pada beberapa hal yaitu:

1. Data base dokumen yang digunakan berjumlah 7 dokumen.
2. Algoritma text mining yang digunakan adalah metode tf-idf.
3. Sistem berbasis web, baru berjalan pada *localhost*.

### 1.5 Metode Penelitian

Metode penelitian yang digunakan diantaranya:

1. Teknik Pengumpulan data

Pengumpulan data terdiri dari data studi pustaka dan observasi. Kepustakaan terdiri dari buku baik *text book* maupun *e-book* dan juga pada penelitian-penelitian terkait sebelumnya. Sedangkan untuk data penelitian

2. Teknik Pengolahan data

Pengolahan data menggunakan metode *Text Mining*, melalui tahapan *stemming*, *stopword removal*, dan *tokenizing*. Dilanjutkan dengan menghitung *Query term* dari dokumen yang ada pada *data base*, lalu melihat jumlah vektor terbesar dari hasil perhitungan. Dokumen yang paling mendekati kesesuaian dengan pencarian adalah dokumen yang memiliki nilai vektor terbesar. Sehingga dokumen tersebutlah yang kemudian dijadikan sebagai rekomendasi.

3. Analisis dan Perancangan Sistem

Analisis dan perancangan sistem menggunakan metode *waterfall*, meliputi tahapan analisis, desain, penyusunan *script* sistem, kemudian dilakukan test.

## II. TINJAUAN PUSTAKA

## 2.1 Sistem Temu Kembali Informasi

Temu kembali Informasi yaitu berkaitan dengan representasi, pencarian dan manipulasi kumpulan besar teks elektronik dan data bahasa lainnya. (Cormack, 2010)

Sistem temu kembali informasi yang dirancang khusus untuk tujuan menemukan informasi yang tersimpan didalam sistem komputer disebut search engine. Tujuan mesin pencari adalah untuk mengurangi jumlah waktu yang di perlukan untuk menemukan informasi yang berguna dari sumber data, dan untuk menyediakan sebuah antarmuka yang dapat menyembunyikan detail mesin yang rumit dari pengguna (user). (Whissel, 2009)

Sistem temu kembali informasi terdiri dari tiga komponen utama, yaitu masukan (input), pemroses (processor) dan keluaran (output). Inti dari sistem temu kembali informasi adalah mencari dokumen-dokumen yang relevan sesuai dengan masukan (query) dari pengguna.

## 2.2 Text Mining

Text mining dapat diartikan sebagai proses pencarian intensif dari pengguna (User) yang berinteraksi dengan dokumen berbeda dari waktu ke waktu dengan menggunakan rangkaian alat analisis.

Text mining berusaha mengekstrak informasi yang bermanfaat dari sumber data melalui identifikasi dan eksplorasi pola menarik

Dalam *text mining* terdapat terdapat istilah *preprocessing* data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data *numerik*. Pada proses *preprocessing* dilakukan beberapa tahap untuk dapat siap diproses memasuki tahap *text mining*. (Nurdiana, Jumadi, & Nursantika, 2016)

Tahap-tahap tersebut antara lain adalah sebagai berikut :

- *Tokenizing / Parsing*

Proses yang dilakukan seseorang untuk menjadikan sebuah kalimat menjadi lebih bermakna atau berada dengan cara memecah kalimat tersebut menjadi kata-kata atau *frase-frase*

- *Stopword*

Proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi, apakah termasuk di dalam daftar kata tidak penting (*stoplist*) atau tidak. Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan di *remove* dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata penting atau *keywords*.

- *Stemming*

Mengurangi ukuran dari suatu ukuran index file. Misalnya dalam suatu deskripsi terdapat *variant* kata “memberikan”, “diberikan”, “memberi”, dan “diberi” hanya memiliki akar kata (*stem*) yaitu “beri”. Ukuran file daftar *index* yang semula berjumlah lima *record* akan di-*reduce* sehingga menjadi satu *record* saja.

- *Tagging*

Tahap mencari bentuk awal atau root dari tiap kata lampau atau kata nhasil *stemming* yang bukan bahasa indonesia.

- *Analyzing*

Tahap penentuan seberapa jauh keterhubungan antara kata-kata dengan dokumen yang ada

## 2.3 Pembobotan TF IDF

Metode Tf-Idf adalah penggabungan dua konsep untuk perhitungan bobot suatu kata (term) yaitu frekuensi kemunculan sebuah kata didalam dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata tersebut menunjukkan seberapa penting dan seberapa umum kata tersebut di dalam dokumen. (Intan & Defeng, 2006)

Berikut di bawah ini adalah rumus umum untuk metode TF-Idf.

$$W_{ij} = tf \times idf \quad (1)$$

$$W_{tj} = tf_j \times \log \frac{N}{n} \quad (2)$$

Dimana :

$W_{ij}$  = bobot kata / term ( $t_j$ ) terhadap dokumen ( $d_i$ )

$tf_{ij}$  = jumlah kemunculan term ( $t_j$ ) dalam dokumen ( $d_i$ )

$N$  = jumlah semua dokumen yang ada dalam database

$n$  = jumlah dokumen yang mengandung term ( $t_j$ ) (minimal ada satu kata yaitu term ( $t_j$ )).

Berapapun besarnya nilai  $tf_{ij}$ , apabila  $N = n$ , maka akan didapatkan hasil 0 (nol), hal itu dikarenakan hasil dari log 1. Maka untuk perhitungan IDF, dapat ditambahkan nilai 1 pada sisi IDF, sehingga perhitungan bobotnya menjadi sebagai berikut :

$$W_{ij} = tf_{ij} \times \log \frac{N}{n} + 1 \quad (3)$$

Rumus di atas dapat dinormalisasi dengan tujuan untuk menstandarisasi nilai bobot ke dalam interval 0 s.d 1. Seperti pada rumus di bawah ini :

$$W_{ij} = \frac{tf_{ij} \times (\log \frac{N}{n} + 1)}{\sqrt{\sum_{k=1}^t (tf_{ik})^2 \times [(\log \frac{N}{n} + 1)]^2}} \quad (4)$$

#### 2.4 Cosine Similarity

Metode Cosine Similarity merupakan metode yang digunakan untuk menghitung similarity (tingkat kesamaan) antar dua buah objek yang dinyatakan dalam dua buah vector dengan menggunakan keywords dari sebuah dokumen sebagai ukuran. (Nurdiana, Jumadi, & Nursantika, 2016)

$$CosSim(d, q) = \frac{q \cdot d}{|q| |d|} = \frac{\sum_{j=1}^t (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij})^2} \sqrt{\sum_{j=1}^t (d_{ij})^2}} \quad (5)$$

Keterangan :

$q_{ij}$  = bobot istilah  $j$  pada dokumen  $i = tf_{ij} \cdot idf_j$

$d_{ij}$  = bobot istilah  $j$  pada dokumen  $i = tf_{ij} \cdot idf_j$

#### 2.5 Confusion Matrix

*Confusion matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. (Indriani, 2014)

Tabel 1. Confusion matrix untuk klasifikasi biner

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan untuk Tabel 1 di atas dinyatakan sebagai berikut :

*True Possitive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.

*True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar dan diklasifikasikan sebagai kelas 0.

*False Possitive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah dan diklasifikasikan sebagai kelas 1.

*False Negative* (FN), yaitu jumlah dokumen dari kelas 1 yang salah dan diklasifikasikan sebagai kelas 0.

Perhitungan akurasi dinyatakan dalam persamaan berikut di bawah ini. (Paskianti, 2011)

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (6)$$

### III. METODE

Pada penelitian ini menggunakan metode deskriptif, yaitu metode yang menggambarkan suatu keadaan atau permasalahan yang sedang terjadi berdasarkan fakta dan data-data yang diperoleh dan dikumpulkan pada waktu melaksanakan penelitian.

Berikut adalah penjelasan terperinci mengenai sistematika model metode penelitian agar lebih mudah dipahami.

#### 3.1 Data dan Informasi yang Diperlukan

Data atau informasi yang diperoleh dalam penelitian ini, dapat digolongkan sebagai berikut :

- a. Data Primer  
Data primer adalah data yang diperoleh dari peninjauan secara langsung ke tempat wisata yang menjadi objek penelitian.
- b. Data Sekunder  
Data sekunder adalah data yang diperoleh melalui studi kepustakaan dengan mempelajari buku dan jurnal yang berhubungan dengan topik dan data yang berkaitan dengan penelitian.

#### 3.2 Sumber Data dan Informasi

- a. Person  
Yakni orang yang menjadi narasumber guna memperoleh data dan informasi yang kemudian berkonsultasi mengenai masalah yang berkaitan dengan penelitian yaitu pengelola tempat wisata.
- b. Paper  
Merupakan sumber tertulis yang berisi data dan informasi yakni dengan cara membaca buku maupun jurnal yang berkaitan dengan penelitian
- c. Place  
Merupakan tempat penelitian yang menjadi sumber untuk memperoleh data dan informasi yaitu Dinas Pariwisata Kab. Subang.

#### 3.3 Teknik Pengumpulan Data

Pada tahap ini dilakukan teknik pengumpulan data dan informasi adalah sebagai berikut :

- a. Observasi  
Pengumpulan data dengan melakukan pengamatan atau meninjau secara langsung terhadap objek wisata yang menjadi bahan penelitian kemudian mencatat hal-hal penting yang berhubungan penelitian ini.
- b. Studi Pustaka  
Yaitu teknik pengumpulan data untuk memperoleh data sekunder dengan menggunakan informasi dari buku maupun jurnal dan sumber lainnya seperti internet yang berkaitan dengan penelitian ini.

#### 3.4 Analisa TF-Idf

Langkah selanjutnya adalah lakukan analisa terhadap metode TF-IDF yaitu mempelajari rumus-rumus dan mulai menghubungkan data yang diperoleh dengan rumus-rumus tersebut sehingga data tersebut menjadi data baru yang akan menghasilkan perhitungan terhadap temu balik informasi lokasi wisata di kabupaten Subang.

#### 3.5 Implementasi (implementasi)

Setelah data dianalisa kemudian lakukan pelatihan menggunakan dua tahapan yakni secara manual dan secara *software*. Tahapan yang dilakukan adalah sebagai berikut :

- Implementasi Secara Manual  
Dilakukan dengan cara mengolah data yang telah diperoleh dari proses analisa secara manual yang kemudian data tersebut di hitung menggunakan rumus perhitungan metode TF-IDF dan lakukan perankingan menggunakan metode *Cosine Similarity*
- Implementasi Secara *Software*

Data yang telah di analisa akan dilakukan pelatihan secara software, menggunakan PHP dan database MySQL. Data yang digunakan harus akurat agar dapat digunakan untuk menganalisa sistem temu balik informasi lokasi wisata di kabupaten Subang.

### 3.6 Pengujian

Pada tahap ini dilakukan untuk mengetahui cara sistem dalam mengolah data. Proses pengujian berfokus pada logika internal software, memastikan bahwa semua pernyataan sudah diuji, dan proses pengujian juga berfokus pada eksternal fungsional, yaitu mengarahkan pengujian untuk menemukan kesalahan-kesalahan dan memastikan bahwa data yang diinputkan akan memberikan hasil aktual yang sesuai dengan hasil yang dibutuhkan.

## IV. HASIL DAN PEMBAHASAN

Pada penelitian ini akan dirancang sebuah form pencarian menggunakan metode TF-IDF. Dalam merancang sebuah sistem kemiripan dokumen, hal pertama yang dilakukan adalah menghitung bobot kata setiap dokumen.

Namun sebelum melakukan pembobotan kata, perlu dilakukan proses *Preprocessing* terlebih dahulu terhadap setiap dokumen termasuk *query* yang dalam hal ini query dimasukan dengan dokumen.

### 4.1 Analisa Pembobotan Kata Menggunakan Metode TF IDF

Dalam penelitian ini peneliti mengasumsikan terdapat 7 dokumen wisata termasuk query. Tabel 1 di bawah ini merupakan contoh isi dokumen 1 yang peneliti asumsikan pada sistem pencarian dokumen wisata.

Tabel 2. Sistem pencarian dokumen wisata

NO DOKUMEN	ISI DOKUMEN
Q	Mencari lokasi wisata bahari di Kabupaten Subang
D1	Tangkuban Perahu terletak di perbawisataan subang dan kabupaten bandung barat yang meyajikan pemandangan yang menakjubkan yaitu kawah yang bisa dinikmati dengan jarak yang sangat dekat dengan suhu yang dingini cocok untuk bersantai dan refleksi dari penat hiruk pikuk curug debu asap panas dll.Dengan ketinggian 2084 m MDPL menyajika pemandangan curug subang yang sangat indah. untuk mencapai ke lokasi pihak pengelola memfasiliwisatai akses jalan yang mudah dan tak lupa juga pemandangan perkebunan teh ketika dalam perjalanannya. Capolaga wisata alam ini menawarkan keindahan ekosistem sungai Cimuja yang unik dan mempesona yang berupa air terjun Cimuja dan air terjun lainnya. Akan tetapi capolaga mempunyai perbedaan yaitu dengan goa - goa dengan nama goanya yaitu goa badak serta memberikan ruang untuk para pencinta alam untuk camp di lokasi curug ini. Curug Cibareubeuy Curug ini terletak di desa Cibeusi, Kecamatan Ciater. Tingginya sekitar 70 meter membelah tebing batu gunung Karamat. Alirannya terlihat berkelok mengikuti kontur tebing, membuat curug ini terlihat unik. Tumpahan airnya membentuk kolam kecil di bawah curug. Di kolam inilah para wisatawan biasanya bermain-main air sepuasnya. Puas bermain air anda bisa beristirahat di saung – saung yang berada tak jauh dari curug. Saung – saung bambu beratap ijuk ini di buat oleh warga setempat bernama pak Ocid. Dirinya kemudian menamakannya Kampung Senyum. Ia bersama istrinya membangun Kampung Senyum untuk wisatawan di lahan kebunnya yang berada tak jauh dari lokasi curug. Saung – saung ini juga dilengkapi toilet dan mushola.Curug Cimuja Curug ini terletak di desa Cibeusi, Kecamatan Ciater. Tingginya sekitar 70 meter membelah tebing batu gunung Karamat. Alirannya terlihat berkelok mengikuti kontur tebing, membuat curug ini terlihat unik. Tumpahan airnya membentuk kolam kecil di bawah curu. Curug Cijalu air terjun dengan ketinggian 70 meter ini terletak di camping ground hutan pinus dalam satu lokasi terdapat 2 curug lainnya yang lebih kecil yaitu curug putri dan curug cilempar. Curug Cileat Ketinggian curug yang mencapai 100 m ini menjadikan percikan air membias dalam warna-warna serta kealiamian tempat ini yang sulit dijamak bagi masyarakat yang berlibur tidak seperti curug lainnya. air panas ciater wisata pemandian panas yang bersumber dari tangkubah perahu dan juga banyak wahana bermain keluarga, cafe,hotel,rest area, oleh - oleh dan juga pacuan kuda.

Tahap selanjutnya adalah melakukan tahap *tokenizing* dan tahap *stopword* pada proses *preprocessing*. Pada tahap ini mengubah semua kata yang mengandung huruf kapital menjadi huruf biasa ( huruf kecil ), menghilangkan semua karakter selain huruf a sampai huruf z, memecah kata sehingga menjadi kata yang berdiri sendiri dan kemudian melakukan *stopword removal* yaitu menghapus kata yang dianggap kurang berarti. Berikut Tabel 2 di bawah ini menunjukkan isi dokumen D1 setelah proses *tokenizing* dan *stopword removal*.

Tabel 3. Dokumen setelah proses *tokenizing* dan *stopword removal*

NO DOKUMEN	ISI DOKUMEN
Q	mencari lokasi wisata bahari di kabupaten subang
D1	tangkuban perahu terletak di perbawisataan subang dan kabupaten bandung barat yang menyajikan pemandangan yang menakjubkan yaitu kawah yang bisa dinikmati dengan jarak yang sangat dekat dengan suhu yang dingini cocok untuk bersantai dan refleksi dari penat hiruk pikuk curug debu asap panas dll dengan ketinggian 2084 m mdpl menyajika pemandangan curug subang yang sangat indah untuk mencapai ke lokasi pihak pengelola memfasiliwisatai akses jalan yang mudah dan tak lupa juga pemandangan perkebunan teh ketika dalam perjalanannya. capolaga wisata alam ini menawarka keindahan ekosistem sungai cimuja yang unik dan mempesona yang berupa air terjun cimuja dan air terjun lainnya akan tetapi capolaga mempunyai perbedaan yaitu dengan goa goa dengan nama goanya yaitu goa badak serta memberikan ruang untuk para pencinta alam untuk camp di lokasi curug ini curug cibareubeuy curug ini terletak di desa cibeusi kecamatan ciater tingginya sekitar 70 meter membelah tebing batu gunung karamat alirannya terlihat berkelok mengikuti kontur tebing membuat curug ini terlihat unik tumpahan airnya membentuk kolam kecil di bawah curug di kolam inilah para wisatawan biasanya bermain-main air sepuasnya puas bermain air anda bisa beristirahat di saung saung yang berada tak jauh dari curug saung saung bambu beratap ijuk ini di buat oleh warga setempat bernama pak ocid dirinya kemudian menamakannya kampung senyum ia bersama istrinya membangun kampung senyum untuk wisatawan di lahan kebunnya yang berada tak jauh dari lokasi curug. saung saung ini juga dilengkapi toilet dan mushola curug cimuja curug ini terletak di desa cibeusi kecamatan ciater tingginya sekitar 70 meter membelah tebing batu gunung karamat alirannya terlihat berkelok mengikuti kontur tebing membuat curug ini terlihat unik tumpahan airnya membentuk kolam kecil di bawah curug curug cijalu air terjun dengan ketinggian 70 meter ini terletak di camping ground hutan pinus dalam satu lokasi terdapat 2 curug lainnya yang lebih kecil yaitu curug putri dan curug cilempar curug cileat ketinggian curug yang mencapai 100 m ini menjadikan percikan air membias dalam warna-warna serta kealamian tempat ini yang sulit dijamak bagi masyarakat yang berlibur tidak seperti curug lainnya. air panas ciater wisata pemandian panas yang bersumber dari tangkuban perahu dan juga banyak wahana bermain keluarga café hotel rest area oleh oleh dan juga pacuan kuda.

Langkah selanjutnya adalah lakukan tahap *stemming* yakni tahap ketiga pada proses *preprocessing*. Yang dilakukan pada tahap ini adalah mengubah kata yang mengandung imbuhan menjadi kata dasar.

Tabel 4. Dokumen setelah proses *stemming*

NO DOKUMEN	ISI DOKUMEN
Q	cari lokasi wisata bahari kabupaten subang
D1	tangkuban perahu letak bawisata subang kabupaten bandung barat yang saji pandang takjub kawah nikmat jarak dekat suhu dingin cocok santai refleksi penat hiruk pikuk curug debu asap panas tinggi 2084 m mdpl saji pandang curug subang indah capai lokasi pihak kelola fasiliwisata akses jalan mudah lupa pandang kebun teh jalan. capolaga wisata alam tawar indah ekosistem sungai cimuja unik pesona rupa air terjun cimuja air terjun lain capolaga punya beda goa goa nama goa goa badak serta beri ruang para cinta alam camp lokasi curug curug cibareubeuy curug letak desa cibeusi kecamatan ciater tinggi sekitar 70 meter belah tebing batu gunung karamat aliran lihat kelok ikuti kontur tebing buat curug lihat unik tumpah air bentuk kolam kecil bawah curug kolam para wisata biasa main main air puas puas main air istirahat saung saung ada jauh curug saung saung bambu atap ijuk buat oleh warga setempat nama pak ocid diri nama kampung senyum sama istri bangun kampung senyum wisata lahan kebun ada jauh lokasi curug. saung saung lengkap toilet mushola curug cimuja curug letak desa cibeusi kecamatan ciater tinggi sekitar 70 meter belah tebing batu gunung karamat aliran lihat kelok ikut kontur tebing buat curug lihat unik tumpah air bentuk kolam kecil bawah curug curug cijalu air terjun tinggi 70 meter letak camping ground hutan pinus satu lokasi dapat 2 curug lain lebih kecil curug putri curug cilempar curug cileat tinggi curug capai 100 m jadi percik air bias warna warna serta alami tempat sulit jamah bagi masyarakat libur curug lain. air panas ciater wisata mandi panas sumber tangkuban perahu banyak wahana main keluarga café hotel rest area oleh oleh pacuan kuda.

Proses selanjutnya adalah lakukan seleksi fitur, yaitu mencari nilai TF-IDF.

Tabel 5. Proses perhitungan TF dan DF pencarian dokumen wisata

No	Term	Q	D1	D2	D3	D7	Df
1	Cari	1	0	0	0	0	1
2	Lokasi	1	4	1	1	1	6
3	Wisata	1	2	0	6	0	5

No	Term	Q	D1	D2	D3	D7	Df
4	Bahari	1	0	0	0	0	1
5	kabupaten	1	1	0	1	0	4
6	Subang	1	2	0	2	0	5
7	Tangkuban	0	2	0	0	0	1
8	Perahu	0	2	0	0	0	1
9	Bawisata	0	1	1	0	0	2
10	bandung	0	1	0	0	0	1
11	Barat	0	1	0	0	0	1
12	Saji	0	2	0	0	0	2
13	pandang	0	3	1	3	1	4
14	Takjub	0	1	0	0	1	2
15	Kawah	0	1	0	0	0	1

Kemudian lakukan perhitungan untuk menghitung nilai IDF dan TF IDF. Nilai TF IDF yang telah dihitung akan digunakan selanjutnya untuk menghitung tingkat kemiripan (similarity) antara *query* dengan setiap dokumen.

Berikut Tabel 5 di bawah ini merupakan contoh perhitungan untuk dokumen 1.

Tabel 6. Nilai TF-IDF pencarian dokumen wisata

Idf	Wdt = tf*idf	
log(n/df)	Q	D1
0,845098	0,84509804	0
0,066947	0,06694679	0,267787159
0,146128	0,14612804	0,292256071
0,845098	0,84509804	0
0,243038	0,24303805	0,243038049
0,146128	0,14612804	0,292256071
0,845098	0	1,69019608
0,845098	0	1,69019608
0,544068	0	0,544068044
0,845098	0	0,84509804
0,845098	0	0,84509804
0,544068	0	1,088136089
0,243038	0	0,729114146
0,544068	0	0,544068044
0,845098	0	0,84509804

Setelah melakukan proses pembobotan kata, proses selanjutnya adalah melakukan perankingan atau kemiripan ( *Similarity*) query terhadap setiap dokumen dengan menggunakan metode *Cosine Similarity*.

#### 4.2 Perankingan Dokumen Menggunakan Metode *Cosine Similarity*

Langkah pertama yang dilakukan adalah lakukan hitungan perkalian antara nilai *wdq* dan *wdi*. Di bawah ini Tabel 6. merupakan perhitungan *wdq\*wdi*.

Tabel 7. *wdq\*wdi* pencarian dokumen wisata

No	Wdq*Wdi				
	D1	D2	D3	...	D7
1	0	0	0	...	0
2	0,017927	0,004481	0,004481		0,004481
3	0,042706	0	0,128120		0
4	0	0	0		0

No	Wdq*Wdi				
	D1	D2	D3	...	D7
5	0,059067	0	0,059067		0
6	0,042706	0	0,042706		0
7	0	0	0		0
8	0	0	0		0
9	0	0	0		0
10	0	0	0		0
11	0	0	0		0
12	0	0	0		0
13	0	0	0		0
14	0	0	0		0
15	0	0	0		0

Langkah kedua yang dilakukan adalah menentukan panjang vektor. Seperti terlihat pada Tabel 7 dibawah ini merupakan contoh perhitungan panjang vektor untuk dokumen 1 ( D1).

Tabel 8. Panjang vektor pencarian dokumen wisata (D1)

NO	Panjang Vektor	
	Q	D1
1	0,714190697	0
2	0,004481873	0,071709962
3	0,021353403	0,085413611
4	0,714190697	0
5	0,059067493	0,059067493
6	0,021353403	0,085413611
7	0	2,856762789
8	0	2,856762789
9	0	0,296010037
10	0	0,714190697
11	0	0,714190697
12	0	1,184040148
13	0	0,531607438
14	0	0,296010037
15	0	0,714190697

Kemudian selanjutnya langkah ketiga yang dilakukan adalah menghitung jumlah total hasil dari perkalian  $wdq*wdi$  dan menghitung jumlah total dari panjang vektor pencarian dokumen wisata.

Tabel 9. Total hasil perkalian  $wdq*wdi$  pencarian dokumen wisata

No	wdq*wdi				
	D1	D2	D3	..	D7
1	0,162408595	0,004481873	0,234376588	...	0,004481873

Tabel 10. Total panjang vektor pencarian dokumen wisata

No	Panjang Vektor				
	D1	D2	D3	...	D7
1	1,534637566	356,7562063	290,9780109	...	86,76258157



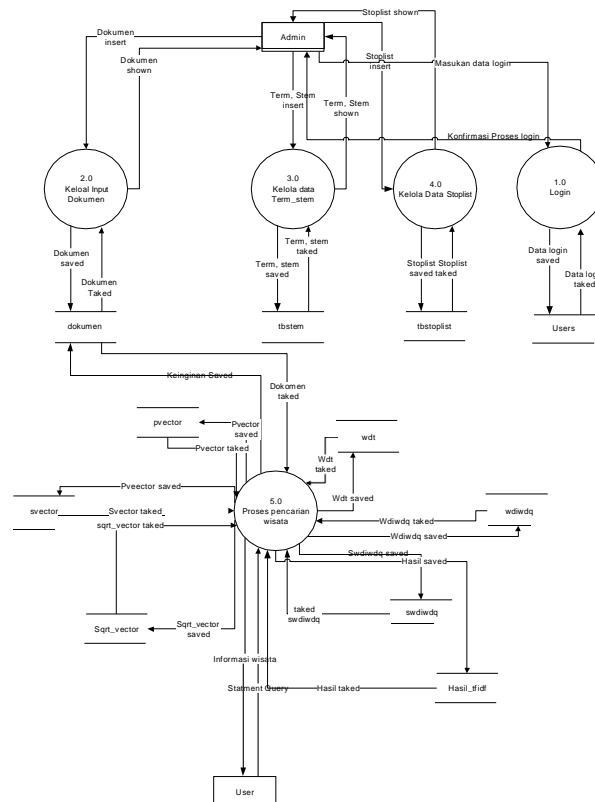


Gambar 2. Diagram konteks sistem pencarian dokumen wisata

#### 4.5 DFD

*Data Flow Diagram* merupakan diagram yang digunakan untuk memodelkan sistem secara logik, yang dapat digunakan pada tahap analisis maupun tahap desain.

Diagram Level 1 Memberikan pandangan secara menyeluruh mengenai sistem yang ditangani, menunjukkan tentang fungsi-fungsi utama atau proses yang ada, aliran data, dan eksternal entity. Data flow diagram level 1 sistem pencarian dokumen wisata, seperti yang tergambar pada Gambar 3 di bawah ini:



Gambar 3. Data Flow Diagram ( DFD ) Level 1

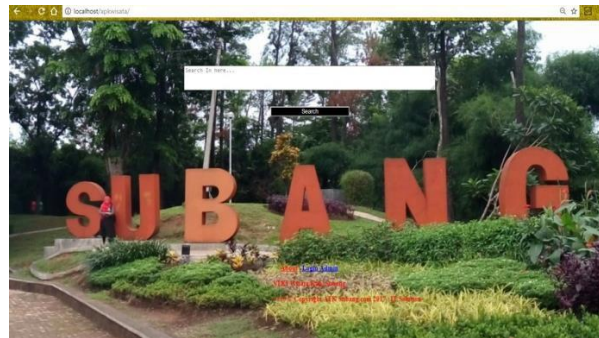
#### 4.6 Implementasi

Aplikasi penentuan dokumen wisata ini menggunakan xampp sebagai database dan menggunakan notepad++ sebagai editor untuk merancang tampilan antar muka. Untuk lebih jelas mengenai tampilan antar muka dapat dilihat pada penjelasan dibawah ini :

##### a. Tampilan Menu Utama User

Pada tampilan menu utama user ini, user di minta untuk memasukkan informasi mengenai pencarian tempat wisata Kabupaten Subang yang kemudian akan diproses oleh sistem dan selanjutnya sistem akan menampilkan hasil pencarian.

Berikut Gambar 4 di bawah ini merupakan tampilan menu utama user.



Gambar 4. Tampilan menu utama user

b. Form Hasil Temu Kembali Informasi Pencarian Wisata

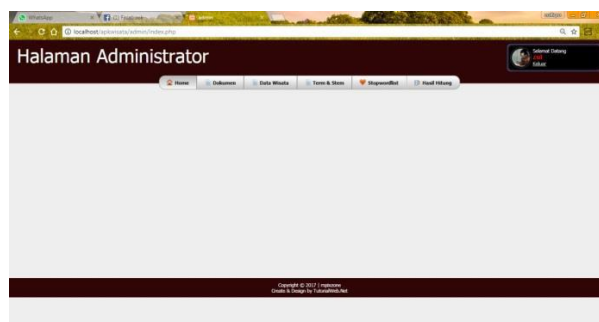
Tampilan form pada Gambar 5 di bawah ini merupakan form hasil temu kembali informasi pencarian wisata Kabupaten Subang berdasarkan informasi yang telah dimasukkan oleh user pada halaman utama.



Gambar 5. Tampilan form hasil temu kembali informasi pencarian wisata

c. Halaman Utama Admin

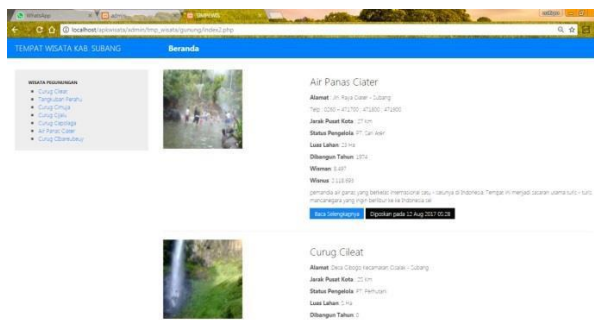
Pada halaman utama admin terdapat 6 tampilan link dalam pencarian dokumen wisata yakni link home, dokumen, data wisata, term & stem, stopwordlist, dan link hasil hitung. Berikut tampilan halaman utama admin.



Gambar 6. Tampilan halaman utama admin

d. Tampilan menu dokumen

Tampilan halaman menu dokumen pada Gambar 7 di bawah ini merupakan tampilan untuk melihat data dokumen pencarian wisata Kabupaten Subang. Pada halaman ini memuat informasi mengenai nama tempat wisata, profil tempat wisata, alamat tempat wisata, jarak dari pusat kota, status pengelola, luas lahan, wisman (wisatawan nusantara), dan wisnus (wisatawan nusantara).



Gambar 7. Tampilan menu dokumen

#### 4.7 Pengujian

Untuk pengujian tingkat akurasi dalam sistem temu kembali pencarian lokasi wisata kabupaten Subang berbasis web menggunakan metode *tf-idf* dan *cosSim* adalah menggunakan *Confusion matrix*.

Tabel 13. Hasil data uji untuk pengujian tingkat akurasi

Confusion Matrik			
Data	Actual	Predict	Ket
Data 1	Dok 1	Dok 1	Relevant
Data 2	Dok 2	Dok 2	Relevant
Data 3	Dok 3	Dok 3	Relevant
Data 4	Dok 4	Dok 4	Relevant
Data 5	Dok 5	Dok 5	Relevant
Data 6	Dok 6	Dok 6	Relevant
Data 7	Dok 7	Dok 7	Relevant
Data 8	Dok 2	Dok 2	Relevant
Data 9	Dok 3	Dok 3	Relevant
Data 10	Dok 2	Dok 2	Relevant
Data 11	Dok 1	Dok 1	Relevant
Data 12	Dok 1	Dok 1	Relevant
Data 13	Dok 3	Dok 3	Relevant
Data 14	Dok 6	Dok 6	Relevant
Data 15	Dok 7	Dok 7	Relevant
Data 16	0	0	Relevant
Data 17	Dok 7	Dok 3	Irrelevant
Data 18	Dok 3	Dok 3	Relevant
Data 19	Dok 5	Dok 5	Relevant
Data 20	Dok 4	Dok 4	Relevant
Data 21	Dok 4	Dok 4	Relevant
Data 22	Dok 4	Dok 4	Irrelevant
Data 23	Dok 4	Dok 4	Relevant
Data 24	Dok 2	Dok 2	Relevant
Data 25	Dok 2	Dok 2	Relevant
Data 26	Dok 4	Dok 4	Relevant
Data 27	Dok 1	Dok 1	Relevant

Confusion Matrik			
Data	Actual	Predict	Ket
Data 28	Dok 1	Dok 6	Irrelevant
Data 29	Dok 5	Dok 5	Relevant
Data 30	Dok 1	Dok 1	Relevant

Tabel 14. Hasil perhitungan mencari nilai data relevan dan tidak relevan.

		Kelas Prediksi	
		Relavan	Tidak Relavan
Kelas Sebenarnya	Relevan	27	0
	Tidak Relevan	0	3

Berdasarkan Tabel 14 di atas, langkah selanjutnya adalah menghitung nilai akurasi menggunakan persamaan (6).

$$Akurasi = \frac{relevan}{tidak relevan} \times 100\%$$

$$Akurasi = \frac{27}{3} \times 100\% = 90\%$$

## V. PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan dapat diambil beberapa kesimpulan, sebagai berikut :

1. Dengan menggunakan teknologi LBS (*Location Based Service*) pengguna dapat mengetahui lokasi wisata di Kota Kabupaten Subang.
2. Dalam menu pencarian, pengguna dapat memasukkan kata kunci berdasarkan statement bebas sehingga pengguna dapat menentukan wisata yang akan dituju.
3. Penerapan perhitungan menggunakan metode TF-IDF berbasis PHP dalam system temu kembali informasi pencarian wisata di Kabupaten Subang memudahkan dalam penentuan bobot kata dan mendapatkan kesimpulan sesuai pencarian yang dituliskan oleh user.
4. Proses pencarian dilakukan dengan cara menghitung bobot kata dan bobot kalimat tiap dokumen yang di inputkan.

### 5.2 Saran

Dari hasil analisis kesimpulan di atas, maka untuk penelitian selanjutnya kami sarankan beberapa hal diantaranya:

1. Upgrade sistem selanjutnya menggunakan teknologi *Location Base Service* (LBS).
2. Pengolahan dokumen tidak hanya menggunakan bahasa indonesia, namun juga bisa menggunakan bahasa inggris agar wisatawan asing dapat menggunakan aplikasi.
3. Pada penelitian selajutnya bisa dilakukan optimasi waktu pemrosesan algoritma TF-IDF ataupun melakukan komparasi antara dua atau lebih algoritma metode *Text Mining*.

## DAFTAR PUSTAKA

- Cormack, S. B. (2010). *Implementing and Evaluating Search Engines*. USA: massachusetts Institute of Technology.
- Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook*. New York: Cambridge University Press.

- Indriani, A. (2014, Juni 21). Klasifikasi Data Forum dengan Menggunakan Metode Naive Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 5-10. Retrieved Juni 25, 2017, from <https://media.neliti.com/publications>
- Intan, R., & Defeng, A. (2006, Juni). HARD: Subject-Based Search Engine Menggunakan TF-IDF Dan Jaccard's Coefficient. *Jurnal Teknik Industri*, 8, 61-72. Retrieved Juni 20, 2017, from <http://www.petra.ac.id/~puslit/journals/dir.php?DepartementID=IND>
- Nurdiana, O., Jumadi, & Nursantika, D. (2016, Juni). Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemahan Al-Qur'an Dalam Bahasa Indonesia. *JOIN*, 1, 59-63. Retrieved Juli 20, 2017, from [id.portalgaruda.org](http://id.portalgaruda.org)
- Paskianti, K. (2011, Juli). *Klasifikasi Dokumen Tumbuhan Obat Menggunakan Algoritma KNN Fuzzy*. Institut Pertanian Bogor, Departemen Ilmu Komputer, Bogor. Retrieved Juli 10, 2017, from [https://docplayer.info/58889956-klasifikasi-dokumen-tumbuhan-obat-menggunakan-algoritma-knn-fuzzy-kristina-paskianti.html#download\\_tab\\_content](https://docplayer.info/58889956-klasifikasi-dokumen-tumbuhan-obat-menggunakan-algoritma-knn-fuzzy-kristina-paskianti.html#download_tab_content)
- W, S. A., Kusyanti, A., & Data, M. (2011). *Modul 3 Database Entity Relationship Diagram*. Malang: Brawijaya University. Retrieved Maret 18, 2018, from <http://power.lecture.ub.ac.id/files/2015/03/Modul-Basis-Data-I-3-ERD.pdf>
- Whissel, J. (2009). *Information Retrieval Using Lucene and Wordnet*, 6.