

AUTOMATIC TEXT  
SUMMARIZATION DOKUMEN  
BAHASA DAERAH  
MENGUNAKAN METODE  
TERMFREQUENCY –INVERSE  
DOCUMENT FREQUENCY(TF-  
IDF) BERBASIS WEB

*by Ahmad Izzuddin*

---

**Submission date:** 01-Apr-2023 08:45PM (UTC+0500)

**Submission ID:** 2052887227

**File name:** document\_1.pdf (502.25K)

**Word count:** 3315

**Character count:** 20624



## *AUTOMATIC TEXT SUMMARIZATION*

### **DOKUMEN BAHASA DAERAH MENGGUNAKAN METODE TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY (TF-IDF) BERBASIS WEB**

#### *AUTOMATIC TEXT SUMMARY*

### *REGIONAL LANGUAGE DOCUMENTS USING THE TERM METHOD FREQUENCY – INVERSE DOCUMENT FREQUENCY (TF-IDF) WEB BASED*

<sup>6</sup> <sup>1</sup> Ahmad Izzuddin, <sup>2</sup> Moch. Ghufron, <sup>3</sup> Imam Marzuki

<sup>1</sup>Mahasiswa Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga

<sup>2</sup>Dosen Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga

<sup>3</sup>Dosen Program Studi Teknik Elektro, Fakultas Teknik, Universitas Panca Marga

e-mail: [izzuddin@upm.ac.id](mailto:izzuddin@upm.ac.id)<sup>1</sup>, [ghufrontx@gmail.com](mailto:ghufrontx@gmail.com)<sup>2</sup>, [imam@upm.ac.id](mailto:imam@upm.ac.id)<sup>3</sup>

#### **Abstract**

One way to preserve the local language is to compose and read local language texts. To understand a text as a whole will be very tiring. If there is a summary result of a text or document, it will be able to help readers to get the essence of a text and can speed up reading time, for that an automatic summary system needs to be done because summarizing manually takes quite a lot of time. In this study the method used was TF-IDF, in this method calculations were carried out by weighing each word in a sentence. So the word that appears the most is an important sentence, then it will produce a collection of texts that contain important sentences from the entire text. From the results of system testing that has been carried out by giving a questionnaire to a linguist, produces an average accuracy value of 20% for Madurese language short story documents at a compression level of 10%, 20%, 30%, 40% and 50%, because the contents of the short story is a narrative sentence. In a narrative document, each sentence has an interconnected meaning, if several sentences are omitted from a story, the storyline will not be good.

**Keywords:** Automatic Summarization, TF-IDF, Weighting, Compression.

#### **Abstrak**

Salah satu cara untuk melestarikan bahasa daerah ialah membuat dan membaca teks berbahasa daerah. Untuk memahami suatu bacaan keseluruhan akan sangat melelahkan. Apabila terdapat hasil ringkasan dari sebuah teks atau dokumen, maka akan dapat membantu pembaca untuk mendapatkan intisari dari sebuah teks dan dapat mempercepat waktu membaca, untuk itu sistem peringkasan otomatis perlu dilakukan karena meringkas secara manual membutuhkan waktu yang cukup banyak. Pada penelitian ini metode yang digunakan adalah TF-IDF, pada metode ini dilakukan perhitungan dengan cara membobot setiap kata dalam kalimat. Sehingga kata yang kemunculan paling banyak adalah sebuah kalimat yang penting, selanjutnya akan menghasilkan kumpulan teks yang berisi kalimat penting dari keseluruhan teks. Dari hasil pengujian sistem yang telah dilakukan dengan memberikan kuesioner kepada seorang ahli bahasa, menghasilkan nilai rata-rata akurasi sebesar 20% untuk dokumen cerpen bahasa Madura pada dengan tingkat kompresi 10%, 20%, 30%, 40% dan 50%, karena isi dari cerpen merupakan kalimat naratif. Dalam dokumen naratif setiap kalimat memiliki makna yang saling berhubungan, apabila beberapa kalimat dihilangkan dari sebuah cerita maka jalan cerita menjadi tidak bagus.

**Keywords:** Peringkasan Otomatis, TF-IDF, Pembobotan, Kompresi.

#### **1. Pendahuluan**

Indonesia dikenal sebagai negara multietnik karena Indonesia sangat luas, yang di dalamnya terdapat bermacam-macam suku, ras dan bahasa yang berbeda. Saat ini, tidak sedikit generasi muda yang berasal dari suku Jawa dan Madura tidak memahami bahasanya, khususnya bahasa yang paling halus (tingkatan ketiga). Hal ini disebabkan karena beberapa faktor, antara lain adalah jarang masyarakat yang menggunakan bahasa daerahnya, khususnya anak-anak muda, lebih-lebih masyarakat yang ada di perkotaan, mereka lebih suka mempergunakan bahasa "gaul" dari pada mempergunakan bahasa daerahnya. Bahasa Indonesia sebagai bahasa nasional justru

Ahmad Izzuddin

Jurnal ENERGY (Jurnal Ilmiah Ilmu-ilmu Teknik) Vol. 10 No. 1 (2020)

<sup>1</sup> menurunkan pamor atau peran bahasa daerah sebagai bahasa etnik [9]. Tidak sedikit anggota masyarakat di daerah peneliti mulai dari ibu-ibu muda, anak-anak hingga remaja menggunakan bahasa Indonesia sebagai bahasa keseharian.

Bahasa Jawa merupakan bahasa daerah terbesar pertama dari 10 besar bahasa daerah di Indonesia, sekitar 84,3 juta jiwa jumlah penuturnya, sedangkan bahasa Madura merupakan bahasa daerah terbesar ketiga, sekitar 13,6 juta jiwa [4]. Sayang sekali apabila bahasa etnik tersebut semakin lama semakin tidak digunakan dan berakhir nah. Salah satu cara untuk melestarikan bahasa daerah ialah membuat dan membaca teks berbahasa daerah [14]. Bagi mereka yang belum paham dan pemula tentang bahasa daerah, membaca keseluruhan dokumen untuk memahaminya akan sangat melelahkan. Apabila ada hasil ringkasan dari suatu teks tersebut, maka akan dit membantu pembaca untuk mendapatkan intisari dari sebuah bacaan dan mempercepat waktu membaca [14]. Oleh karena itu, peringkasan otomatis perlu dilakukan.

Penelitian mengenai metode peringkasan otomatis telah dilakukan beberapa penelitian terdahulu. Pertama, penelitian [14] yang berusaha membangun aplikasi peringkasan otomatis teks bahasa Jawa menggunakan metode TF-IDF. Kedua, penelitian [6] yang berusaha membandingkan 2 metode yang telah difokuskan TF-IDF menggunakan perhitungan secara global sedangkan Intersection menggunakan perhitungan secara lokal dalam peringkasan dokumen bahasa Indonesia. Ketiga, penelitian [13] yang berusaha meringkas dokumen menggunakan bahasa Bali.

Dari ketiga penelitian di atas, aplikasi yang dibangun tidak satupun memiliki fitur pemilihan peringkasan bahasa lainnya, hanya difokuskan satu bahasa saja. Peringkasan dokumen bahasa daerah yang telah dibangun oleh beberapa penelitian sebelumnya adalah bahasa Jawa dan Bali, namun belum pernah ada peringkasan dokumen bahasa Madura, yang tentunya bahasa Madura sendiri memiliki karakteristik berbeda dengan kebanyakan bahasa lainnya. Untuk itu peneliti akan membangun aplikasi peringkasan multi bahasa yang terdiri dari bahasa Indonesia, Jawa dan Madura. Dalam penelitian ini, peneliti akan mengajukan metode TF-IDF dalam peringkasan dokumen bahasa daerah karena metode TF-IDF merupakan metode ekstraktif. Metode ekstraktif akan memilih secara otomatis kalimat-kalimat yang relevan yang kemudian dapat merepresentasikan keseluruhan isi dari suatu dokumen [6]. Jadi peneliti bermaksud akan mengajukan judul "Automatic Text Summarization Dokumen Bahasa Daerah Menggunakan Metode *Term Frequency – Inverse Document Frequency (TF-IDF)*".

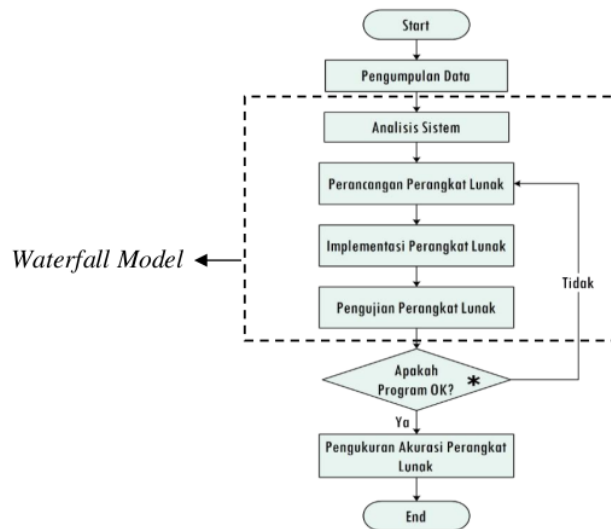
## 2. LANDASAN TEORI

Penelitian yang dilakukan oleh Widiantoro (2014) yang berjudul "Peringkasan teks otomatis pada dokumen berbahasa Jawa menggunakan metode TF-IDF". Pada penelitian ini membahas sistem peringkasan dokumen yang dapat meringkas teks secara otomatis pada teks bahasa Jawa menggunakan metode TF-IDF, dan agar menghasilkan ringkasan dari teks bahasa Jawa yang sesuai dengan kebutuhan pengguna. Penelitian yang dilakukan oleh Pradipta (2016) yang berjudul "Eksplorasi Metode Peringkasan Ekstraktif Untuk Dokumen Ilmiah Bidang Teknologi Informasi". Pada penelitian ini membahas yang berusaha membandingkan 2 metode yang telah difokuskan TF-IDF menggunakan perhitungan secara global dan Intersection menggunakan perhitungan secara lokal. Penelitian yang dilakukan oleh Suputra (2017) yang berjudul "Peringkasan Teks Otomatis Untuk Dokumen Bahasa Bali Berbasis Metode Ekstraktif". Pada penelitian ini membahas sistem peringkasan otomatis yang berbasis metode ekstraktif untuk dokumen berbahasa Bali

Terkait penelitian terdahulu, Aplikasi peringkasan dokumen yang telah dibangun sebelumnya adalah peringkasan dokumen bahasa Jawa, peringkasan dokumen bahasa Bali dan peringkasan dokumen bahasa Indonesia, akan tetapi belum pernah ada penelitian peringkasan teks bahasa daerah, lebih-lebih dokumen bahasa Madura yang masih belum pernah ada peneliti yang mengangkat peringkasan dokumen bahasa Madura. Penelitian selanjutnya akan mengangkat bahasa Madura, bahasa Jawa dan Bahasa Indonesia dalam sistem peringkasan dokumen multi bahasa menggunakan metode TF-IDF.

## 3. Metodologi

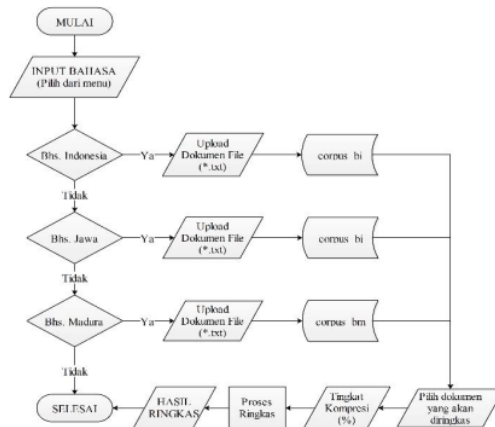
<sup>24</sup> Pada penelitian ini, model perancangan sistem yang digunakan adalah waterfall model yang didalamnya terdapat beberapa tahapan. Adapun tahapan-tahapan tersebut digambarkan oleh Gambar 2



Gambar 3.1. Kerangka penelitian <sup>8</sup>

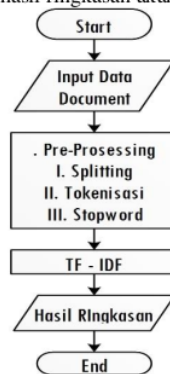
Berdasarkan gambar diatas dapat dijelaskan kerangka penelitian sebagai berikut :

1. **Pengumpulan Data**  
Data yang digunakan peneliti pada penelitian ini ada dua. Pertama data primer: dengan melakukan observasi <sup>5</sup> dan yang kedua data sekunder: data yang diperoleh dari jurnal dan internet.
2. **Analisis Sistem**  
Pada tahap ini peneliti melakukan analisis sistem untuk dijadikan bahan referensi <sup>22</sup> da penelitian sebelumnya, guna sebagai acuan untuk merancang dan membangun sesuai dengan kebutuhan sistem.
3. **Perancangan Perangkat Lunak**  
Pada tahap ini peneliti melakukan perancangan sistem peringkasan teks bahasa daerah yang akan dibangun, <sup>16</sup> r memperoleh gambaran secara detail.
4. **Implementasi Perangkat Lunak**  
Implementasi perangkat lunak dilakukan dengan membuat sistem peringkasan dokumen multi bahasa yang dapat melas <sup>9</sup> kan ringkasan kepada pengguna dan mampu mendapatkan intisari dari sebuah dokumen
5. **Pengujian Perangkat Lunak**  
Pada tahap ini peneliti melakukan pengujian sistem peringkasan dokumen otomatis untuk menemukan kekurangan dari sistem tersebut.
  - A. **Perancangan <sup>9</sup> perangkat Lunak**  
Pada tahap ini peneliti melakukan perancangan sistem peringkasan teks bahasa daerah yang akan dibangun. Berikut flowchart perancangan sistem sistem peringkasan dokumen :



Gambar 3.2. Diagram alir sistem

Flowchart sistem pada peringkasan dokumen multi bahasa, user memulai aplikasi, kemudian input bahasa atau pilih bahasa pada menu bar. Setelah memilih menu bahasa, maka upload dokumen-dokumen yang ingin diringkas terlebih dahulu dengan file extension (\*.txt) kemudian hasil upload tersimpan di dalam storage pada folder corpus tiap bahasa. Lalu pilih dokumen yang ingin diringkas, input tingkat kompresi ringkas sesuai kebutuhan dan klik button ringkas, kemudian hasil ringkasan akan ditampilkan.



Gambar 3.3. Diagram alir peringkasan dokumen otomatis

Berdasarkan gambar diatas berikut langkah-langkah membangun sistem peringkasan otomatis:

- 1) Input Data Dokumen, yaitu Pengambilan dokumen yang ada dalam folder corpus\_bi untuk dokumen bahasa Indonesia, corpus\_bj untuk dokumen bahasa Jawa dan corpus\_bm untuk dokumen bahasa Madura.
- 2) Pre-processing, yaitu sebagai berikut :
  - a. Splitting bertujuan untuk memecah teks yang utuh menjadi pecahan-pecahan kalimat.
  - b. Tokenisasi, bertujuan untuk menghilangkan tanda baca, kata penghubung dan imbuhan.
  - c. Stopword, bertujuan menghilangkan kata (term) yang tidak memiliki nilai informasi.
- 3) TF-IDF, yaitu perhitungan bobot setiap kata dan dokumen yang kemudian nilai bobot kalimat paling rendah akan dihilangkan, sedangkan nilai bobot tertinggi akan diprioritaskan dalam peringkasan dokumen.

19

#### B. Pengujian Perangkat Lunak

Pengujian perangkat lunak ini dilakukan dengan dua cara yaitu :

- 1) Pengujian menggunakan metode Blackbox Testing
- 2) Pengujian menggunakan kuesioner, yaitu mengkoreksi secara manual hasil ringkasan yang diperoleh dari sistem kepada seseorang ahli bahasa atau yang sudah mengerti dibidang meringkas teks atau dokumen.

## 4. Hasil dan Pembahasan 4

### A. Implementasi Sistem

Implementasi sistem yang dilakukan pada penelitian ini terdiri dari dua tahapan utama yaitu tahapan pelatihan dan tahapan pengujian dimana masing-masing tahap memiliki nilai yang saling berhubungan

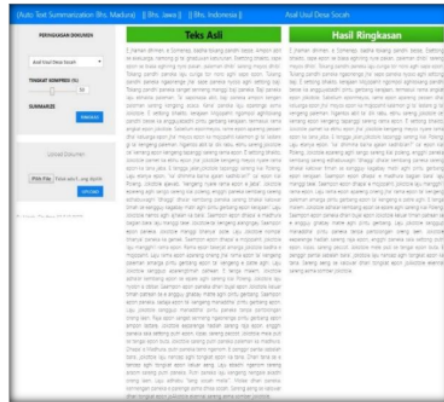
- 1) Tahap Pengunggahan Dokumen  
Pengunggahan dilakukan oleh sistem dengan cara klik tombol unggah pada fitur *upload* dokumen yang nantinya berada pada folder corpus.
- 2) Tahap Pembacaan Dokumen  
Dalam proses pembacaan data ini diambil secara keseluruhan dokumen atau data set pada folder corpus yang telah diunggah sebelumnya.
- 3) Tahap Membuat Array Inverted Index  
12uannya dari membuat *Array Inverted Index* adalah memisahkan kata yang berbeda dengan cara memotong tiap kata (*term*) yang berbeda dari suatu daftar *term* dokumen.
- 4) Tahap Pemanggilan Dokumen  
Dalam proses pemanggilan data ini diambil 1 dokumen atau data set.
- 5) Tahap Mengukur Tingkat Kompresi Ringkasan  
Tujuan untuk tingkat kompresi ringkasan adalah untuk mengukur berapa persen (%) dokumen yang ingin diringkas.
- 6) Tahap Pre – Processing  
Pada tahap ini memecah dokumen menjadi kalimat yang disebut dengan *Splitting*. Memecah kalimat pada suatu teks menjadi kalimat-kalimat dengan cara mendeteksi keberadaan tanda titik “.”, tanda tanya “?” dan tanda tanya “!” sebagai tanda *delimiter* untuk memotong *string* teks atau dokumen.
- 7) Tahap Pembobotan  
Pada proses pembobotan ini terdapat tahapan-tahapan. Tahapan pertama hitung frekuensi/nilai setiap token/kata (TF) pada setiap kalimat. Kemudian hitung nilai *Inversed Document Frequency* (IDF), setelah ditemukan nilai keduanya, maka cari nilai *weight* (W) dengan cara menjumlah nilai TF dan IDF.
- 8) Tahap Penggabungan Kalimat  
Dalam Proses penggabungan kalimat *sorting* bobot tertinggi terbelah dahulu kemudian potong *array* dan urutkan kalimat berdasarkan urutan dokumen aslinya. Setelah itu gabungkan kalimat hasil ringkasan.

### B. Implementasi Antarmuka

Pada tahapan ini, rancangan desain antarmuka yang sebelumnya dibuat akan diubah ke dalam bahasa pemrograman yaitu PHP.



Gambar 4.1. Antarmuka Menu Utama



Gambar 4.2. Antarmuka Aplikasi Dijalankan

Diketahui hasil ringkasan dokumen bahasa Madura gambar di atas dengan judul “Asal Usul Desa Socah” terdapat jumlah 43 kalimat yang akan diringkaskan dengan tingkat kompresi sebanyak 50%. Berikut rumus perhitungannya 50% dari 43 kalimat yaitu :  $43 \times 50\% = 21.5 \approx 22$  Dokumen. Jadi 21.5 dibulatkan menjadi 22 kalimat.

C. Pengujian Perangkat Lunak

Proses evaluasi dengan menggunakan 15 data cerpen bahasa Madura dan Jawa yang diambil pada tanggal 18 Desember 2018. Kemudian hasil ringkasan dikoreksi dan dievaluasi sesuai kusioner yang telah diberikan kepada seorang ahli bahasa Dekan dari Fakultas Sastra & Filsafat Universitas Panca Marga Probolinggo.

Pertanyaan : Apakah hasil ringkasan sudah menggambarkan ide pokok dari suatu dokumen atau sesuai kebutuhan atau keinginan pengguna dengan tingkat kompresi \* ?

Tabel 4.1. Kuesioner Pengujian Hasil Ringkasan Bahasa Madura

Dokumen	Pertanyaan	Jawaban				
		A	B	C	D	E
Doc. 1 – 15	* 10%	x	x	x	x	√
Doc. 1 – 15	* 20%	x	x	x	x	√
Doc. 1 – 15	* 30%	x	x	x	x	√
Doc. 1 – 15	* 40%	x	x	x	x	√
Doc. 1 – 15	* 50%	x	x	x	x	√

11  
 Nilai Jawaban :

A = Sangat Baik, B = Baik, C = Cukup Baik, D = Kurang Baik, E = Buruk

Setelah dilakukan pengisian kusioner dalam pengujian sistem, maka selanjutnya dilakukan perhitungan jumlah *texts file* yang relevan. Dari jumlah teks yang relevan akan lakukan perhitungan nilai akurasi yang dihasilkan oleh sistem. Berikut presentase nilai akurasi peringkasan dari semua tingkat kompresi dengan tingkat kompresi 10%, 20%, 30%, 40% dan 50% hasilnya ialah sama :

A =  $(0/15) \times 20\% = 0\%$   
 B =  $(0/15) \times 40\% = 0\%$   
 C =  $(0/15) \times 60\% = 0\%$   
 D =  $(0/15) \times 80\% = 0\%$   
 E =  $(15/15) \times 100\% = 100\%$

Pertanyaan : Apakah hasil ringkasan sudah menggambarkan ide pokok dari suatu dokumen atau sesuai kebutuhan atau keinginan pengguna dengan tingkat kompresi \* ?

Tabel 4.2. Kuesioner Pengujian Hasil Ringkasan Bahasa Indonesia

Dokumen	Pertanyaan	Jawaban				
		A	B	C	D	E
Doc. 1 – 10	* 25%	x	x	x	√	x
Doc. 1 – 10	* 50%	x	√	x	x	x

Presentase Nilai Akurasi Peringkasan Tingkat Kompresi 25%

$$A = (0/10) \times 20 \% = 0\%$$

$$B = (0/10) \times 40 \% = 0\%$$

$$C = (0/10) \times 60 \% = 0\%$$

$$D = (10/10) \times 80 \% = 100\%$$

$$E = (0/10) \times 100 \% = 0\%$$

Presentase Nilai Akurasi Peringkasan Tingkat Kompresi 50%

$$A = (0/10) \times 20 \% = 0\%$$

$$B = (0/10) \times 40 \% = 100\%$$

$$C = (0/10) \times 60 \% = 0\%$$

$$D = (10/10) \times 80 \% = 0\%$$

$$E = (0/10) \times 100 \% = 0\%$$

Berdasarkan hasil kuesioner di atas pada tabel 4.1, maka dari pengujian sistem dengan dokumen berjumlah 15 dokumen cerpen bahasa Madura dan Jawa, bisa dikatakan sistem dalam meringkas cerpen tersebut mendapatkan hasil keseluruhan jawaban “Buruk” dengan rata-rata tingkat akurasi 20% pada keseluruhan tingkat kompresi yang diuji, dikarenakan isi dari kalimat cerpen merupakan kalimat naratif bukan kalimat deskriptif. Pada dokumen yang bersifat naratif tidak bisa diringkas hanya dengan memotong atau menghilangkan beberapa kalimat saja, dalam dokumen naratif sendiri setiap kalimat memiliki makna dan berhubungan satu sama lain pada setiap kalimatnya, apabila beberapa kalimat dihilangkan dari sebuah cerita maka jalan cerita menjadi tidak bagus bahkan bisa dikatakan aneh karena menghilangkan beberapa scene pada sebuah cerita, sedangkan dokumen deskriptif merupakan pemaparan atau penggambaran suatu objek, jadi kemungkinan besar bisa untuk diringkas dengan hanya memotong atau menghilangkan beberapa kalimat.

Karena dokumen dengan kalimat deskriptif bahasa daerah belum ada jadi peneliti menggunakan dokumen bahasa Indonesia. Dan itu terlihat jelas pada tabel 4.2, setelah sistem meringkas dokumen dengan kalimat deskriptif bahasa Indonesia terdapat hasil yang lebih bagus dari pada dokumen sebelumnya yaitu, pada tingkat kompresi 25% memperoleh jawaban keseluruhan “Kurang Baik (D)” dengan rata-rata tingkat akurasi 40%. Sedangkan pada tingkat kompresi 50% memperoleh yang lebih bagus lagi yaitu jawaban keseluruhan “Baik (B)” dengan rata-rata tingkat akurasi 80%.

Pada tingkat kompresi 25% memperoleh akurasi yang lebih kecil dari pada kompresi 50%, hal itu dikarenakan pada tingkat kompresi 25% ada kalimat yang lebih penting dihilangkan oleh sistem yang seharusnya ditampilkan.

## 4. Kesimpulan

### A. Kesimpulan

Kesimpulan pada penelitian ini yaitu, sistem peringkasan teks bahasa daerah ini sudah berjalan dengan baik, akan tetapi setelah dilakukan peringkasan pada sistem ini dengan dokumen cerpen berbahasa Madura memperoleh jawaban keseluruhan “Buruk (E)” dengan rata-rata nilai akurasi 20%, karena isi dari cerpen merupakan kalimat naratif bukan kalimat deskriptif. Dalam dokumen naratif setiap kalimat memiliki makna yang saling berhubungan, apabila beberapa kalimat dihilangkan dari sebuah cerita maka jalan cerita menjadi tidak bagus karena menghilangkan beberapa scene pada sebuah cerita, sedangkan dokumen deskriptif sendiri merupakan pemaparan atau penggambaran suatu objek yang didalamnya merupakan kalimat yang dapat berisi gambaran sifat-sifat benda yang dideskripsikan, jadi kemungkinan besar bisa untuk diringkas hanya dengan memotong beberapa kalimat. Dan itu terbukti setelah sistem meringkas dokumen dengan kalimat deskriptif bahasa Indonesia memperoleh hasil lebih bagus yaitu pada tingkat kompresi 25% memperoleh keseluruhan jawaban “Kurang Baik (D)” dengan rata-rata nilai akurasi 40%. Sedangkan pada tingkat kompresi 50% memperoleh jawaban “Baik (B)” dengan rata-rata nilai akurasi 80%.



14

#### B. Saran

Adapun saran pada penelitian ini adalah sebagai berikut :

1. Mencari metode alternatif lainnya untuk meringkas dokumen bahasa daerah, menimbang metode TF-IDF tidak bisa meringkas cerpen bahasa daerah.
2. Pada teks dengan kalimat yang panjang diperlukan perlakuan khusus atau mencari suatu metode yang dapat mengatasi teks dengan kalimat yang panjang, menimbang metode yang peneliti gunakan ini hanya menjumlah bobot kata pada suatu kalimat.
3. Penggunaan metode *stemming* dalam tahapan *pre-processing*.
4. Kata yang mengandung singkatan-singkatan lebih baik ditulis lengkap, karena akan mempengaruhi relevansi ringkasan dari sistem.

#### Referensi

- [1] Bakhel. 2019. Pengertian, Contoh, Kelebihan Dan Kekurangan Website, URL: <https://www.bakhel.com/>.
- [2] Diantara, E. 2017. Perancangan Sistem Informasi Monitoring Error Checking Berbasis Web Pada PT Cahaya Lestari Sejahtera. Jurusan Sistem Informasi STMIK Raharja Tangerang.
- [3] Fadilah, Nisa and Hernawati, Kuswari. 2016. Pengembangan Sistem Pengolahan Hasil Belajar Siswa Smp Berbasis Kurikulum 2013 Studi Kasus: SMP Negeri 1 Prambanan. Universitas Negeri Yogyakarta
- [4] Iliyyan, D.U. 2015. Cerpen Bahasa Madura, URL:<http://aksaranai.blogspot.com/>.
- [5] Koran Sindo. 2017. Ini Dia, 10 Bahasa Daerah dengan Penutur Terbanyak di Indonesia, URL:<https://nasional.sindonews.com/>
- [6] Masitoh, D. 2016. Sistem Pendukung Keputusan Seleksi Penerimaan Karyawan Baru Nabila *Cake & Bakery* Dengan Metode *Simple Additive Weighting* (SAW) Berbasis Web. Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Ponorogo.
- [7] Pradipta, Y.A. 2016. Eksplorasi Metode Peringkasan Ekstraktif Untuk Dokumen Ilmiah Bidang Teknologi Informasi. Program Studi S1 Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Maranatha Bandung.
- [8] Rahayu, S.L. 2016. *Peringkasan Teks Otomatis Pada Jurnal Bahasa Indonesia dengan Menggunakan Metode Vector Space Model*. UIN Sunan Gunung Djati Bandung.
- [9] Rahilah., Solihin, F., Rachman, F.H. 2013. Terjemahan Bahasa Madura, URL: <http://www.terjemahan.madura.web.id/>.
- [10] Raihany, A. 2015. Pergeseran Penggunaan Bahasa Madura di Kalangan Anak- anak Sekolah Dasar Negeri di Desa Pangarangan Kecamatan Kota Sumenep. Program Studi Prodi Tadris Bahasa Inggris Jurusan Tarbiyah STAIN Pamekasan.
- [11] Setiawan, P., Sulistiowati., Lemantara, J. 2015. Rancang Bangun Aplikasi Pengolahan Data Evaluasi Proses Belajar Mengajar Berbasis *Web* Pada Stikes Yayasan Rs. Dr. Soetomo Surabaya. Jurusan Sistem Informasi STMIK STIKOM Surabaya.
- [12] Soetedja, I. 2017. Terjemahan Jawa. URL:<https://mongosilakan.net/>.
- [13] Soetomo. 2015. Cerita Rakyat, URL:<http://www.panjebarsemangat.co.id/>.
- [14] Suputra, I.P.G.H. 2017. Peringkasan Teks Otomatis Untuk Dokumen Bahasa Bali Berbasis Metode Ekstraktif. Program Studi Teknik Informatika, Jurusan Ilmu Komputer F'akultas Matematika Dan Ilmu Pengetahuan Alam Universitas Udayana.
- [15] Widianoro, A. 2014. Peringkasan Teks Otomatis Pada Dokumen Berbahasa Jawa Menggunakan Metode TF-IDF. Informatics Engineering Study Program Faculty Of Science And Technology Sanata Dharma University Yogyakarta.

# AUTOMATIC TEXT SUMMARIZATION DOKUMEN BAHASA DAERAH MENGGUNAKAN METODE TERMFREQUENCY – INVERSE DOCUMENT FREQUENCY(TF-IDF) BERBASIS WEB

## ORIGINALITY REPORT

15%

SIMILARITY INDEX

13%

INTERNET SOURCES

5%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://ejournal.iainmadura.ac.id">ejournal.iainmadura.ac.id</a> Internet Source	2%
2	Regina Hoo, Elias R. Ruiz-Morales, Iva Kelava, Carmen Sancho-Serra et al. "Early infection response of the first trimester human placenta at single-cell scale", Cold Spring Harbor Laboratory, 2023 Publication	2%
3	<a href="http://repository.widyatama.ac.id">repository.widyatama.ac.id</a> Internet Source	1%
4	<a href="http://ejournal.itn.ac.id">ejournal.itn.ac.id</a> Internet Source	1%
5	<a href="http://repository.usd.ac.id">repository.usd.ac.id</a> Internet Source	1%
6	<a href="http://jurnal.um-palembang.ac.id">jurnal.um-palembang.ac.id</a> Internet Source	1%
7	<a href="http://download.garuda.ristekdikti.go.id">download.garuda.ristekdikti.go.id</a> Internet Source	1%

8	<a href="http://id.123dok.com">id.123dok.com</a> Internet Source	1 %
9	<a href="http://ejournal.its.ac.id">ejournal.its.ac.id</a> Internet Source	<1 %
10	<a href="http://inacl.id">inacl.id</a> Internet Source	<1 %
11	<a href="http://anzdoc.com">anzdoc.com</a> Internet Source	<1 %
12	<a href="http://repository.usu.ac.id">repository.usu.ac.id</a> Internet Source	<1 %
13	<a href="http://core.ac.uk">core.ac.uk</a> Internet Source	<1 %
14	<a href="http://repositori.uin-alauddin.ac.id">repositori.uin-alauddin.ac.id</a> Internet Source	<1 %
15	<a href="http://as-wait.icu">as-wait.icu</a> Internet Source	<1 %
16	<a href="http://e-journal.uajy.ac.id">e-journal.uajy.ac.id</a> Internet Source	<1 %
17	<a href="http://repositori.usu.ac.id">repositori.usu.ac.id</a> Internet Source	<1 %
18	<a href="http://www.slideshare.net">www.slideshare.net</a> Internet Source	<1 %
19	<a href="http://eprints.uny.ac.id">eprints.uny.ac.id</a> Internet Source	<1 %

20	<a href="http://ikee.lib.auth.gr">ikee.lib.auth.gr</a> Internet Source	<1 %
21	<a href="http://repositorio.unicamp.br">repositorio.unicamp.br</a> Internet Source	<1 %
22	<a href="http://repository.its.ac.id">repository.its.ac.id</a> Internet Source	<1 %
23	<a href="http://digilib.ulm.ac.id">digilib.ulm.ac.id</a> Internet Source	<1 %
24	<a href="http://fdocuments.es">fdocuments.es</a> Internet Source	<1 %
25	<a href="http://jurnal.unissula.ac.id">jurnal.unissula.ac.id</a> Internet Source	<1 %
26	<a href="http://society.fisip.ubb.ac.id">society.fisip.ubb.ac.id</a> Internet Source	<1 %
27	<a href="http://jurnal.stts.edu">jurnal.stts.edu</a> Internet Source	<1 %

Exclude quotes  On

Exclude matches  < 2 words

Exclude bibliography  On

# AUTOMATIC TEXT SUMMARIZATION DOKUMEN BAHASA DAERAH MENGGUNAKAN METODE TERMFREQUENCY – INVERSE DOCUMENT FREQUENCY(TF-IDF) BERBASIS WEB

---

## GRADEMARK REPORT

---

FINAL GRADE

**/0**

GENERAL COMMENTS

**Instructor**

---

PAGE 1

---

PAGE 2

---

PAGE 3

---

PAGE 4

---

PAGE 5

---

PAGE 6

---

PAGE 7

---

PAGE 8

---