

***AUTOMATIC TEXT SUMMARIZATION PADA SINOPSIS  
LAPORAN KECELAKAAN LALU LINTAS PENERBANGAN  
DI INDONESIA MENGGUNAKAN ALGORITMA *TERM  
FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)****



**OLEH :**

**MUHAMMAD FARHAN HILMAN**

**09011281924148**

**JURUSAN SISTEM KOMPUTER**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2023**

***AUTOMATIC TEXT SUMMARIZATION PADA SINOPSIS  
LAPORAN KECELAKAAN LALU LINTAS PENERBANGAN  
DI INDONESIA MENGGUNAKAN ALGORITMA *TERM  
FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)****

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**



**OLEH :**

**MUHAMMAD FARHAN HILMAN**

**09011281924148**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA**

**2023**

**HALAMAN PENGESAHAN**

***AUTOMATIC TEXT SUMMARIZATION PADA SINOPSIS  
LAPORAN KECELAKAAN LALU LINTAS PENERBANGAN DI  
INDONESIA MENGGUNAKAN ALGORITMA TERM  
FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)***

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**

**Oleh :**

**MUHAMMAD FARHAN HILMAN**

**09011281924148**

**Indralaya, <sup>28</sup> November 2023**

**Ketua Jurusan Sistem Komputer**

**Pembimbing Skripsi**

  
**Dr. Ir. Sukemi, M.T.**

**NIP. 196612032006041001**

  
**Rossi Passarella, M.Eng.**

**NIP. 197806112010121004**

## HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jumat

Tanggal : 6 Oktober 2023

Tim Penguji :

1. Ketua Sidang : Dr. Ir. Sukemi, M.T.



( )

2. Sekretaris Sidang : Muhammad Ali Buchari, S.Kom., M.T.



( )

3. Penguji Sidang : Dr. Firdaus, M.kom.



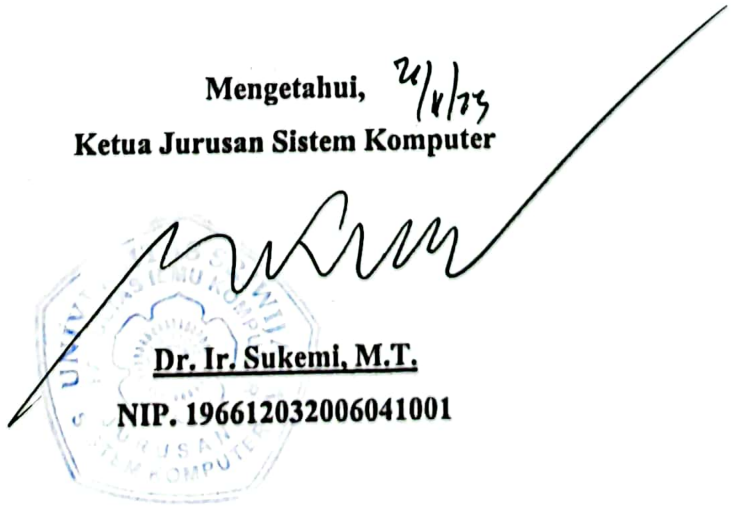
( )

4. Pembimbing : Rossi Passarella, M.Eng.

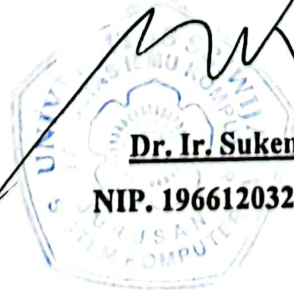


( )

Mengetahui, 20/10/23  
Ketua Jurusan Sistem Komputer



**Dr. Ir. Sukemi, M.T.**  
**NIP. 196612032006041001**



## HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Muhammad Farhan Hilman  
NIM : 09011281924148  
Judul : Automatic Text Summarization pada Sinopsis Laporan  
Kecelakaan Lalu Lintas Penerbangan di Indonesia Menggunakan  
Algoritma Term Frequency-Inverse Document Frequency (TF-  
IDF)

**Hasil Pengecekan Software iThenticate/Turnitin : 14%**

Menyatakan bahwa laporan tugas akhir Saya merupakan hasil karya sendiri dan bukan penjiplakan / *plagiat*. Apabila ditemukan unsur penjiplakan / *plagiat* dalam laporan tugas akhir ini, maka Saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini Saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Indralaya, November 2023



Muhammad Farhan Hilman

09011281924148

## **HALAMAN PERSEMBAHAN**

### **MOTTO**

**“MUNDUR SELANGKAH UNTUK MAJU”**

**“Jangan merasan tertinggal, setiap orang mempunyai proses dan rezekinya masing-masing”**

**(QS Maryam : 4)**

**“Sebaik-baik manusia adalah yang paling bermanfaat bagi orang lain.”**

**-Rasulullah SAW-**

Saya sangat berterima kasih atas doa-doa tulus yang senantiasa orang tua panjatkan, serta perjuangan serta pengorbananmu yang tak terhitung dalam membawa kebahagiaan kepada anak-anakmu. Setiap keringat yang kau tuangkan dan usaha yang kau lakukan adalah wujud cinta yang tak terhingga. Semoga Allah memberikan balasan berlipat ganda atas segala kebaikanmu, dan kami berharap menjadi anak yang membuatmu bangga dan menjadi sumber kebahagiaanmu.

Semoga rahmat Allah senantiasa melimpah bagimu, memberikan kesehatan dan umur yang panjang serta berkah, menjadi cahaya penerang di setiap langkah perjalanan anak-anakmu, menguatkan dalam setiap tantangan, dan mewujudkan kebahagiaan dalam setiap pencapaian. Dalam doa tulus, kami memohon agar kasih sayang-Mu senantiasa melindungi dan menjaga mereka sepanjang hidup.

**Aamiin yaa robbal 'aalamiin.**

## KATA PENGANTAR

Assalammu'alaikum Warrahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin. Puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah memberikan karunia dan rahmat-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini yang berjudul "***Automatic Text Summarization*** pada **Sinopsis Laporan Kecelakaan Lalu Lintas Penerbangan Menggunakan Algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)***".

Dalam proposal ini penulis menjelaskan mengenai penerapan Algoritma TF-IDF pada automatic text summarization untuk melakukan ringkasan pada synopsis laporan KNKT. Dalam penelitian ini text summarization didasarkan pada keyword extractor untuk menemukan kata kunci dan kalimat informatif dengan disertai data-data yang diperoleh penulis saat melakukan penelitian dan pengujian data. Penulis berharap agar tulisan ini dapat bermanfaat bagi banyak orang dan menjadi bahan acuan bagi yang tertarik untuk meneliti tentang penerapan automatic text summarization menggunakan Algoritma *TF-IDF* dalam meringkas teks.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada beberapa pihak atas ide dan saran serta bantuannya dalam menyelesaikan penulisan skripsi ini. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT, yang telah memberikan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penulisan skripsi ini dengan baik dan lancar.
2. Orang tua saya tercinta untuk segala do'a, motivasi dan dukungannya baik moril, materil, maupun spiritual selama ini.
3. Bapak Prof.Dr.Erwin.SSi, MSi, selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. H. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya serta selaku Pembimbing Akademik penulis di Jurusan Sistem Komputer.
5. Bapak Rossi Passarella, S.T. M.Eng., selaku Dosen Pembimbing skripsi yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran

dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan skripsi ini.

6. Bapak Ahmad Fali Oklilas, M.T., selaku dosen pembimbing akademik selama saya menjadi mahasiswa di Universitas Sriwijaya.
7. Mbak Renny selaku admin Jurusan Sistem Komputer yang telah membantu mengurus seluruh berkas.
8. Teman-teman Kinderjoi Family Rakha, Epai, Jakak, Ariba, Namon, Tepek, Zahra, Dilak, Nauk dan Vidra selaku teman seperjuangan yang selalu memberikan semangat dari awal hingga akhir perkuliahan.
9. Teman-teman Fakultas Ilmu Komputer yang sudah memberikan semangat dan motivasi kepada penulis dalam menyelesaikan skripsi ini.
10. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa laporan ini masih sangat jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangatlah diharapkan penulis agar dapat segera memperbaiki, sehingga laporan ini dapat dijadikan sebagai masukan ide dan pemikiran yang bermanfaat bagi semua pihak dan menjadi tambahan bahan bacaan bagi yang tertarik untuk meneliti tentang penerapan *text Summarization* berdasarkan algoritma *TF-IDF*. Akhir kata penulis berharap, semoga skripsi ini bermanfaat dan berguna bagi khalayak.

Wassalammu'alaikum Warahmatullahi Wabarakatuh

Indralaya,

Penulis,



Muhammad Farhan Hilman

NIM. 09011281924148



***AUTOMATIC TEXT SUMMARIZATION PADA SINOPSIS LAPORAN  
KECELAKAAN LALU LINTAS PENERBANGAN DI INDONESIA  
MENGUNAKAN ALGORITMA TERM FREQUENCY-INVERSE  
DOCUMENT FREQUENCY (TF-IDF)***

**Muhammad Farhan Hilman**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Email : [aanaja112233@gmail.com](mailto:aanaja112233@gmail.com)

**ABSTRAK**

Automatic text summarization merupakan salah satu cabang dalam ilmu natural language processing (NLP) yang memiliki tujuan untuk merepresentasikan suatu teks yang panjang kemudian dikompresi sehingga dapat dibaca dan dipahami dengan mudah oleh pengguna. Penerapan algoritma Term Frequency - Inverse Document Frequency (TF-IDF) dalam automatic text summarization dapat menghitung skor dan bobot dari setiap kalimat dalam dokumen sehingga dapat menemukan kalimat penting dalam suatu teks. Dalam penelitian ini, dilakukan penerapan automatic text summarization menggunakan algoritma TF-IDF pada dataset kumpulan sinopsis final report KNKT kecelakaan lalu lintas penerbangan di Indonesia. Percobaan dilakukan pada 142 data sinopsis kemudian hasil ringkasan dari algoritma TF-IDF dilakukan analisis perbandingan ROUGE yang dibandingkan dengan hasil ringkasan manusia dan hasil ringkasan website (<https://www.scribbr.com/text-summarizer/>). Hasil terbaik skor ROUGE dari perbandingan antara ringkasan TF-IDF dan ringkasan manusia adalah ROUGE-1 0.746, ROUGE-2 0.727 dan ROUGE-L 0.746 dengan rata-rata skor ROUGE-1 0.475, ROUGE-2 0.265 dan ROUGE-L 0.453. Sedangkan hasil terbaik skor ROUGE perbandingan antara ringkasan TF-IDF dan ringkasan website adalah ROUGE-1 0.719, ROUGE-2 0.6 dan ROUGE-L 0.719 dengan rata-rata skor ROUGE-1 0.499, ROUGE-2 0.279 dan ROUGE-L 0.478.

**Kata Kunci** : Automatic text summarization, Natural language processing (NLP), Term Frequency - Inverse Document Frequency (TF-IDF), ROUGE

***AUTOMATIC TEXT SUMMARIZATION ON SYNOPSIS OF AIR TRAFFIC  
ACCIDENT REPORTS IN INDONESIA USING TERM FREQUENCY-  
INVERSE DOCUMENT FREQUENCY (TF-IDF) ALGORITHM***

**Muhammad Farhan Hilman**

Department of Computer Systems, Faculty of Computer Science, Sriwijaya  
University.

Email : [aanaja112233@gmail.com](mailto:aanaja112233@gmail.com)

**ABSTRACT**

Automatic text summarization is one of the branches in natural language processing (NLP) which aims to represent a long text and then compressed so that it can be read and understood easily by users. The application of Term Frequency - Inverse Document Frequency (TF-IDF) algorithm in automatic text summarization can calculate the score and weight of each sentence in the document so that it can find important sentences in a text. In this research, automatic text summarization using TF-IDF algorithm is applied to the dataset of KNKT final report synopsis collection of aviation traffic accidents in Indonesia. The experiment was conducted on 142 synopsis data then the summary results of the TF-IDF algorithm were analyzed for ROUGE comparison compared to human summary results and website summary results (<https://www.scribbr.com/text-summarizer/>). The best ROUGE scores from the comparison between TF-IDF summary and human summary are ROUGE-1 0.746, ROUGE-2 0.727 and ROUGE-L 0.746 with an average score of ROUGE-1 0.475, ROUGE-2 0.265 and ROUGE-L 0.453. While the best ROUGE score results of comparison between TF-IDF summary and website summary are ROUGE-1 0.719, ROUGE-2 0.6 and ROUGE-L 0.719 with an average score of ROUGE-1 0.499, ROUGE-2 0.279 and ROUGE-L 0.478.

**Keywords:** Automatic text summarization, Natural language processing (NLP), Term Frequency - Inverse Document Frequency (TF-IDF), ROUGE

## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b> .....	<b>iii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iv</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>v</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>ABSTRAK</b> .....	<b>ix</b>
<b>ABSTRACT</b> .....	<b>x</b>
<b>DAFTAR ISI</b> .....	<b>xi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
1.6.1. Bagian Awal Skripsi .....	5
1.6.2. Bagian Isi Skripsi .....	5
1.6.3. Bagian Akhir Skripsi.....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
2.1 <i>Automatic Text Summarization</i> .....	7
2.2 <i>Natural language Processing</i> .....	10
2.3 Algoritma <i>TF-IDF (Term Frequency-Inverse Document Frequency)</i> ....	12

2.4	Ekstraksi Fitur .....	13
2.5	<i>Recall-Oriented Understudy for Gisting Evaluation (ROUGE)</i> .....	13
2.6	<i>Text Mining</i> .....	15
2.7	Komite Nasional Keselamatan Transportasi (KNKT) .....	16
2.8	Penelitian Terkait .....	17
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>19</b>
3.1	Pendahuluan .....	19
3.2	Pengumpulan Data .....	20
3.2.1	Jenis Data .....	20
3.2.2	Sumber Data.....	20
3.2.3	Metode Pengumpulan Data.....	20
3.3	Perancangan Algoritma .....	20
3.3.1	Data Teks .....	21
3.3.2	<i>Preprocessing</i> .....	21
3.3.3	TF-IDF (Term Frequency-Inverse Document Frequency) .....	22
3.3.4	Menghitung Nilai Skor Kalimat.....	23
3.3.5	Ringkasan.....	23
3.4	Analisis Hasil .....	23
3.5	Penarikan Kesimpulan.....	24
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>25</b>
4.1	Hasil Penelitian .....	25
4.1.1	Persiapan Dataset .....	25
4.1.2	<i>Pre-processing</i> Dataset .....	26
4.1.3	<i>Term Frequency – Inverse Document Frequency</i> .....	34
4.1.4	Hasil ringkasan.....	37
4.2	Pembahasan.....	38

<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>49</b>
5.1 Kesimpulan .....	49
5.2 Saran.....	50
<b>DAFTAR PUSTAKA .....</b>	<b>51</b>
<b>LAMPIRAN.....</b>	<b>54</b>

## DAFTAR GAMBAR

Gambar 3. 1 <i>Flowchart</i> metodologi penelitian .....	19
Gambar 3. 2 Perancangan algoritma .....	21
Gambar 3. 3 <i>Preprocessing</i> .....	21
Gambar 3. 4 TF-IDF .....	22
Gambar 4. 1 Hasil <i>Preprocessing</i> .....	26
Gambar 4. 2 <i>python code library pre-processing</i> .....	34
Gambar 4. 3 Hasil <i>Preprocessing</i> .....	34
Gambar 4. 4 Perbandingan skor <i>ROUGE-1</i> .....	44
Gambar 4. 5 Perbandingan skor <i>ROUGE-2</i> .....	45
Gambar 4. 6 Perbandingan skor <i>ROUGE-L</i> .....	45

## DAFTAR TABEL

Tabel 2. 1 Hasil evaluasi ROUGE score penelitian terkait .....	18
Tabel 4. 1 Dataset.....	25
Tabel 4. 2 Case Folding .....	27
Tabel 4. 3 <i>Word Tokenization</i> .....	29
Tabel 4. 4 <i>Stopword Removal</i> .....	31
Tabel 4. 5 <i>Stemming</i> .....	32
Tabel 4. 6 Pemisahan Kalimat .....	35
Tabel 4. 7 Skor tiap Kalimat .....	36
Tabel 4. 8 Hasil Ringkasan .....	37
Tabel 4. 9 Perbandingan Ringkasan.....	39
Tabel 4. 10 Hasil Analisis <i>Rouge</i> .....	39
Tabel 4. 11 Skor <i>ROUGE-1</i> Tertinggi .....	40
Tabel 4. 12 Skor <i>ROUGE-2</i> Tertinggi .....	40
Tabel 4. 13 Skor <i>ROUGE-L</i> Tertinggi .....	41
Tabel 4. 14 Skor <i>ROUGE-1</i> Terendah .....	41
Tabel 4. 15 Skor <i>ROUGE-2</i> Terendah .....	42
Tabel 4. 16 Skor <i>ROUGE-L</i> Terendah.....	43
Tabel 4. 17 Skor rata-rata <i>ROUGE</i> .....	43
Tabel 4. 18 Skor <i>ROUGE</i> tertinggi, terendah dan rata-rata .....	46
Tabel 4. 19 Perbandingan R-1 dan R-2 peneliti dengan penelitian terkait .....	47

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Perkembangan teknologi yang terjadi di era globalisasi saat ini berkembang sangat cepat sehingga menuntut setiap orang agar selalu meng-*update* informasi terbaru. Kebutuhan manusia untuk memperoleh informasi saat ini merupakan suatu kebutuhan yang sangat penting. Dalam memperoleh informasi tersebut manusia dapat memanfaatkan berbagai media, seperti halnya media digital. Jumlah dokumen digital yang diterbitkan di *World Wide Web* dan perpustakaan digital telah berkembang pesat dalam beberapa tahun terakhir karena perkembangan teknologi informasi dan komunikasi. Informasi tentang *world wide web* terus bertumbuh dengan tingkat eksponensial [1]. Maka dari itu, dengan berkurang dan singkatnya waktu membaca menjadi hal yang sangat dibutuhkan oleh manusia tanpa melewatkan informasi penting dalam suatu bacaan. Berbagai informasi yang beredar dan berkembang sangat cepat menjadi hal yang membuat manusia sulit dalam memahami berbagai informasi sekaligus. Membaca banyaknya informasi dan berita secara terus-menerus menyebabkan berkurangnya banyak waktu, dan membuat manusia malas karena tidak sesuai dengan keinginan masyarakat yang ingin memperoleh informasi secara cepat, tepat dan akurat. Sedangkan manusia memiliki kecepatan membaca yang rendah, sehingga sulit untuk memahami informasi dan berita dengan cepat. Salah satu cara yang dapat menjadi solusi agar kita bisa memahami berita sekaligus adalah dengan membaca rangkuman atau ringkasan, yang merupakan inti atau bagian penting dalam suatu bacaan. Oleh karena itu, memberikan ringkasan dari suatu informasi tanpa kehilangan signifikansinya menjadi hal yang sangat diperlukan.

*Automatic text summarization* telah dijadikan suatu cara alternatif untuk dapat menyelesaikan masalah-masalah diatas. *Automatic text summarization* merupakan salah satu cabang dalam ilmu *natural language processing (NLP)* yang memiliki tujuan untuk merepresentasikan suatu teks yang panjang kemudian dikompresi sehingga dapat dibaca dan dipahami dengan mudah oleh pengguna [2]. *Text summarization* adalah praktik memecah publikasi panjang menjadi paragraf atau



kalimat yang dapat diatur. Prosedur mengekstrak informasi penting sambil juga memastikan bahwa pengertian paragraf dipertahankan. Ini mempersingkat waktu yang diperlukan untuk memahami materi panjang seperti artikel penelitian tanpa menghilangkan informasi penting. Identifikasi topik, interpretasi, dan pembuatan ringkasan adalah langkah-langkah dalam proses peringkasan teks [3].

*Text summarization* dapat diterapkan dalam dua metode yaitu *abstractive text summarization* dan *extractive text summarization*. Dimana pada *abstractive summarization* dilakukan pengolahan kata atau kalimat pada suatu teks kemudian di abstraksi agar mendapatkan kalimat baru atau secara artian *abstractive summarization* adalah metode peringkasan abstraktif, mencari inti dari keseluruhan teks dan menyajikan arti dari isi teks tersebut. Cara ini dapat menciptakan kata dan frasa, kemudian menyatukannya dengan cara yang bermakna, selain itu juga menambahkan fakta paling penting yang ditemukan dalam teks. Dengan cara ini, teknik yang digunakan lebih kompleks dibanding teknik ekstraktif secara komputasi. Berbeda dengan *abstractive text summarization*, *extractive text summarization* merupakan metode peringkasan teks dengan Teknik atau cara memilih atau mengklasifikasikan antara kata penting dan tidak penting kemudian membuang kata tidak penting agar suatu kalimat lebih ringkas tanpa kehilangan inti dari suatu kalimat. Tujuan utama dari *text summarization* adalah meringkas suatu teks tanpa kehilangan inti penting dalam teks tersebut [4].

Selain itu, *Text summarization* bisa dibagi dalam dua kategori, ialah *multi-document summarization* dan *single-document summarization*, berdasarkan pada jumlah input yang berasal dari dokumen yang diberikan. Daripada itu, *text summarization* diidentifikasi sebagai *query-based* atau *generic text summarization*. *query-based text summarization* merangkum dokumen berdasarkan *query* yang diajukan oleh pengguna, sedangkan *Generic summarization* merangkum banyak konten dari dokumen [5].

Ragam pendekatan konvensional dalam *text summarization* yang berbasis ekstraktif umumnya melibatkan intervensi manusia dalam penerapannya. Misalnya, ini melibatkan gabungan fitur-fitur statistik dan linguistik seperti *term frequency*, membandingkan panjangnya dan posisi kalimat, serta berbagai teknik lain. seperti halnya, metode *SumBasic* dilakukan dengan merangking kalimat menggunakan

frekuensi kata dan memilih kalimat utama sebagai ringkasan. Kemudian metode *Latent semantic analysis (LSA)* diterapkan untuk pengukuran kesamaan antara kalimat-kalimat didasarkan dengan persamaan matriks semantik yang telah dihasilkan. Selain itu, berbagai pendekatan lainnya telah diperkenalkan, seperti peringkasan teks berbasis jaringan saraf (neural), peringkasan teks menggunakan *naïve Bayes*, dan beragam metode lainnya [6].

Metode *Term Frequency - Inverse Document Frequency (TF-IDF)* dapat menghitung skor dan berat dari setiap kalimat dalam dokumen, contoh dalam suatu kasus peringkasan di mana kalimat dengan skor/bobot tertinggi akan dipilih untuk memberikan kesimpulan dari dokumen atau teks. Namun, pilihan berdasarkan data terbesar ini menyebabkan kalimat pertama ringkasan terkadang terbaca tidak nyambung [7]. Hal ini dikarenakan fakta bahwasannya sebagian besar kalimat dengan skor yang tinggi adalah kalimat yang berisikan kata-kata penting, yang dimana pada dokumen penelitian sering digunakan simbol ilmiah tertentu yang dibaca oleh sistem sebagai kata penting.

Dari permasalahan tersebut, peneliti ingin melakukan *percobaan extractive text summarization* pada laporan KNKT menggunakan algoritma *TF-IDF*. Dataset yang akan digunakan dalam penelitian ini adalah sinopsis pada laporan kecelakaan lalu lintas penerbangan di Indonesia yang dapat di unduh secara publik dan dipublikasikan di *website* KNKT sehingga bisa digunakan sebagai data *text summarization* yang peneliti rancang dengan metode lain. Maka dari itu, penelitian ini mengambil judul yaitu : “*Automatic text summarization* pada sinopsis laporan kecelakaan lalu lintas penerbangan di Indonesia menggunakan algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)*”.

## 1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang sebelumnya, maka dapat disimpulkan rumusan masalah dalam penelitian ini adalah:

1. Bagaimana proses penerapan algoritma *TF-IDF* untuk *extractive text summarization*.
2. Bagaimana hasil analisis dan evaluasi kualitas *extractive text summarization* dengan algoritma *TF-IDF*.

### 1.3 Batasan Masalah

Dalam penelitian ini memerlukan batasan-batasan untuk tercapainya tujuan penelitian. Berikut beberapa batasan masalah dalam penelitian ini ialah:

1. Metode yang digunakan dalam penelitian ini adalah *extractive text summarization*.
2. Data yang digunakan untuk penelitian ini adalah kumpulan sinopsis laporan kecelakaan lalu lintas penerbangan di Indonesia.
3. Algoritma *extractive text summarization* yang digunakan adalah Algoritma *TF-IDF*.

### 1.4 Tujuan Penelitian

Adapun beberapa tujuan dalam penelitian ini, antara lain:

1. Mengetahui proses implementasi Algoritma *TF-IDF* dalam *proses extractive text summarization*.
2. Mengetahui kualitas hasil ringkasan yang dihasilkan berdasarkan hasil analisis pada *extractive text summarization* dengan algoritma *TF-IDF*.
3. Menghasilkan *extractive text summarization* yang dapat dijadikan benchmarking terhadap penelitian yang sejenis.

### 1.5 Manfaat Penelitian

Adapun beberapa manfaat dalam penelitian ini adalah sebagai berikut :

1. Memahami penerapan dan implementasi metode Algoritma *TF-IDF* dalam penelitian *extractive text summarization*.
2. Mengetahui hasil analisis kualitas ringkasan yang dihasilkan menggunakan metode Algoritma *TF-IDF*.
3. Memberikan kontribusi terhadap penelitian *extractive text summarization*.
4. Dengan penerapan *extractive text summarization* pada laporan KNKT penerbangan dengan Algoritma *TF-IDF* dapat mengetahui efektifitas ringkasan sinopsis pada laporan tersebut dan menemukan kalimat penting.

## **1.6 Sistematika Penulisan**

Sistematika penulisan ini dirancang dengan tujuan memudahkan pemahaman rangkaian pemikiran yang terkandung dalam skripsi ini. Secara garis besar, penulisan skripsi ini dibagi menjadi tiga bagian, seperti berikut ini.

### **1.6.1. Bagian Awal Skripsi**

Pada bagian awal skripsi terdapat elemen-elemen seperti halaman judul, halaman pernyataan, halaman motto dan persembahan, abstrak terdiri dari halaman judul, halaman pengesahan, pernyataan, motto dan persembahan. Selain itu juga terdapat abstrak, kata pengantar, daftar isi, gambar, tabel serta lampiran.

### **1.6.2. Bagian Isi Skripsi**

Bagian isi skripsi terdiri dari lima bab, yaitu sebagai berikut.

#### **1. BAB 1: PENDAHULUAN**

Bab ini berisikan pembahasan mengenai latar belakang, perumusan masalah, batasan masalah, tujuan serta manfaat dari penelitian ini, serta sistematika penulisan skripsi.

#### **2. BAB 2: TINJAUAN PUSTAKA**

Bab ini berisi penjelasan mengenai definisi dan kerangka teoritis yang memberikan dasar bagi pemahaman serta solusi terhadap permasalahan yang diangkat dalam skripsi ini.

#### **3. BAB 3: METODE PENELITIAN**

Bab ini menguraikan langkah-langkah dalam penelitian, termasuk tahap studi pendahuluan, pemilihan dataset, perancangan algoritma, pengembangan sistem, analisis hasil, dan kesimpulan yang diambil.

#### **4. BAB 4: HASIL DAN PEMBAHASAN**

Bab ini, akan dipaparkan mengenai keseluruhan hasil penelitian bersertakan pembahasannya.

#### **5. BAB 5: PENUTUP**

Bab ini berisikan sebagai tempat menyimpulkan seluruh konten skripsi dan memberikan saran-saran untuk pengembangan lebih lanjut.

### **1.6.3. Bagian Akhir Skripsi**

Bagian akhir skripsi memuat daftar pustaka yang mencatat sumber-sumber, referensi dan buku-buku yang digunakan penulis, serta lampiran-lampiran yang memberikan dukungan bagi isi skripsi ini.

## DAFTAR PUSTAKA

- [1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102264, 2020, doi: 10.1016/j.ipm.2020.102264.
- [2] M. A. Zamzam, "Sistem Automatic Text Summarization Menggunakan Algoritma Textrank," *Matics*, vol. 12, no. 2, pp. 111–116, 2020, doi: 10.18860/mat.v12i2.8372.
- [3] C. Setyawan, N. Benarkah, and V. R. Prasetyo, "Automatic Text Summarization Berdasarkan Pendekatan Statistika pada Dokumen Berbahasa Indonesia," *KELUWIH J. Sains dan Teknol.*, vol. 2, no. 1, pp. 1–7, 2021, doi: 10.24123/saintek.v2i1.4045.
- [4] M. A. Pai, "Text Summarizer Using Abstractive and Extractive Method," *Int. J. Eng. Res. Technol.*, vol. Vol. 3, no. 5, pp. 971–976, 2014, [Online]. Available: <http://www.ijert.org/view-pdf/9752/text-summarizer-using-abstractive-and-extractive-method>.
- [5] N. Hendrastuty and A. SN, "Text Summarization in Multi Document Using Genetic Algorithm," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 4, p. 327, 2021, doi: 10.22146/ijccs.66026.
- [6] K. E. Dewi and N. I. Widiastuti, "The Design of Automatic Summarization of Indonesian Texts Using a Hybrid Approach," *J. Teknol. Inf. dan Pendidik.*, vol. 15, no. 1, pp. 37–43, 2022, doi: 10.24036/jtip.v15i1.451.
- [7] S. Krimberg, N. Vanetik, and M. Litvak, "Summarization of financial documents with TF-IDF weighting of multi-word terms," *Proc. 3rd Financ. Narrat. Process. Work. FNP 2021*, pp. 75–80, 2021.
- [8] P. Studi, T. Informatika, and M. M. Mubarak, *Indonesian Abstractive News Summarization Berbasis Deep Learning Dengan Metode Sequence-.* 2021.
- [9] G. Sharma and D. Sharma, "Automatic Text Summarization Methods: A Comprehensive Review," *SN Comput. Sci.*, vol. 4, no. 1, 2023, doi: 10.1007/s42979-022-01446-w.

- [10] I. W. A. Setyadi, D. C. Khrisne, and I. M. A. Suyadnya, "Automatic Text Summarization Menggunakan Metode Graph dan Metode Ant Colony Optimization," vol. 17, no. 1, pp. 124–130, 2018.
- [11] N. F. Saraswati, Indriati, and R. S. Perdana, "Peringkasan Teks Otomatis Menggunakan Metode Maximum Marginal Relevance Pada Hasil Pencarian Sistem Temu Kembali Informasi Untuk Artikel Berbahasa Indonesia," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 5494–5502, 2018, doi: 10.1016/s1010-6030(01)00380-x.
- [12] D. P. Ismi and F. Ardianto, "Peringkasan Ekstraktif Teks Bahasa Indonesia dengan Pendekatan Unsupervised Menggunakan Metode Clustering," *Cybernetics*, vol. 3, no. 02, p. 90, 2020, doi: 10.29406/cbn.v3i02.2290.
- [13] I. Huda, "Implementasi Natural Language Processing (Nlp) Untuk Aplikasi Pencarian Lokasi," *J. Nas. Teknol. Terap.*, vol. 3, no. 2, p. 15, 2021, doi: 10.22146/jntt.35036.
- [14] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [15] Z. Xu and J. Zhang, "Extracting Keywords from Texts based on Word Frequency and Association Features," *Procedia Comput. Sci.*, vol. 187, pp. 77–82, 2021, doi: 10.1016/j.procs.2021.04.035.
- [16] J. M. Sanchez-Gomez, M. A. Vega-Rodríguez, and C. J. Pérez, "A multi-objective memetic algorithm for query-oriented text summarization: Medicine texts as a case study," *Expert Syst. Appl.*, vol. 198, no. December 2020, p. 116769, 2022, doi: 10.1016/j.eswa.2022.116769.
- [17] R. Adelia, S. Suyanto, and U. N. Wisesty, "Indonesian abstractive text summarization using bidirectional gated recurrent unit," *Procedia Comput. Sci.*, vol. 157, pp. 581–588, 2019, doi: 10.1016/j.procs.2019.09.017.
- [18] R. A. Sinoara, J. Antunes, and S. O. Rezende, "Text mining and semantics: a systematic mapping study," *J. Brazilian Comput. Soc.*, vol. 23, no. 1, 2017, doi: 10.1186/s13173-017-0058-7.

- [19] Komite Nasional Keselamatan Transportasi Reublik Indonesia, “Laporan investigasi kecelakaan lalu lintas penerbangan di Indonesia,” *KNKT.go.id*. <https://knkt.go.id/investigasi> (accessed Sep. 20, 2022).
- [20] Z. Cai, N. Lin, C. Ma, and S. Jiang, “Indonesian automatic text summarization based on a new clustering method in sentence level,” *ACM Int. Conf. Proceeding Ser.*, pp. 30–35, 2019, doi: 10.1145/3341620.3341626.
- [21] G. Garmastewira and M. L. Khodra, “Summarizing Indonesian news articles using Graph Convolutional Network,” *J. Inf. Commun. Technol.*, vol. 18, no. 3, pp. 345–365, 2019, doi: 10.32890/jict2019.18.3.6.
- [22] K. Kurniawan and S. Louvan, “I NDO S UM : A New Benchmark Dataset for Indonesian Text Summarization,” *2018 Int. Conf. Asian Lang. Process.*, pp. 215–220, 2018.