



**KERJA PRAKTIK - IF184801**

**Prediksi Teks Wawancara Medis Berdasarkan Nilai Kemiripan menggunakan Cosine Similarity**

Institut Teknologi Sepuluh Nopember

Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota Surabaya, Jawa Timur 60111

Periode: 1 Oktober 2022 – 14 Mei 2023

**Oleh:**

I Kadek Ricky Suirta

05111840000146

**Pembimbing Jurusan**

Dini Adni Navastara, S.Kom., M.Sc.

**Pembimbing Lapangan**

Dr. Ir. Diana Purwitasari, S.Kom., M.Sc.

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2023



**KERJA PRAKTIK - IF184801**

**Prediksi Teks Wawancara Medis Berdasarkan Nilai Kemiripan menggunakan Cosine Similarity**

Institut Teknologi Sepuluh Nopember  
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota Surabaya, Jawa Timur 60111  
Periode: 1 Oktober 2022 – 14 Mei 2023

Oleh:

I Kadek Ricky Suirta

05111840000146

**Pembimbing Jurusan**

Dini Adni Navastara, S.Kom., M.Sc.

**Pembimbing Lapangan**

Dr. Ir. Diana Purwitasari, S.Kom., M.Sc.

DEPARTEMEN TEKNIK INFORMATIKA  
Fakultas Teknologi Elektro dan Informatika Cerdas  
Institut Teknologi Sepuluh Nopember  
Surabaya 2023

*[Halaman ini sengaja dikosongkan]*

## DAFTAR ISI

<b>DAFTAR ISI</b> .....	iv
<b>DAFTAR GAMBAR</b> .....	ix
<b>DAFTAR KODE SUMBER</b> .....	xiii
<b>LEMBAR PENGESAHAN</b> .....	xv
<b>KATA PENGANTAR</b> .....	xxi
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1. Latar Belakang</b> .....	1
<b>1.2. Tujuan</b> .....	2
<b>1.3. Manfaat</b> .....	2
<b>1.4. Rumusan Masalah</b> .....	2
<b>1.5. Lokasi dan Waktu Kerja Praktik</b> .....	3
<b>1.6. Metodologi Kerja Praktik</b> .....	3
<b>1.6.1. Perumusan Masalah</b> .....	3
<b>1.6.2. Studi Literatur</b> .....	3
<b>1.6.3. Implementasi Sistem</b> .....	3
<b>1.6.4. Pengujian dan Evaluasi</b> .....	3
<b>1.6.5. Kesimpulan dan Saran</b> .....	4
<b>1.7. Sistematika Laporan</b> .....	4
<b>1.7.1. Bab I Pendahuluan</b> .....	4
<b>1.7.2. Bab II Profil Perusahaan</b> .....	4

1.7.3.	Bab III Tinjauan Pustaka.....	4
1.7.4.	Bab IV Metodologi.....	4
1.7.5.	Bab V Pengujian dan Evaluasi.....	4
1.7.6.	Bab VI Kesimpulan dan Saran .....	4
<b>BAB II PROFIL PERUSAHAAN .....</b>		<b>5</b>
2.1.	Profil Departemen Informatika ITS.....	5
2.2.	Lokasi.....	6
2.3.	Visi dan Misi.....	6
2.3.1	Visi.....	6
2.3.2	Misi .....	7
2.3.3	Tujuan.....	7
2.3.4	Sasaran.....	8
2.4.	Laboratorium Bidang Minat .....	8
2.4.1	Lab. Rekayasa Perangkat Lunak (RPL).....	8
2.4.2	Lab. Komputasi Berbasis Jaringan (KBJ).....	9
2.4.3	Lab. Komputasi Cerdas dan Visi (KCV) .....	10
2.4.4	Lab. Arsitektur dan Jaringan Komputer (AJK)	11
2.4.5	Lab. Grafika Interaksi dan Game (GIGa) .....	12
2.4.6	Lab. Algoritma dan Pemrograman (AP) .....	13
2.4.7	Lab. Manajemen Cerdas Informasi (MCI)....	14
2.4.8	Lab. Pemodelan dan Komputasi Terapan (PKT)	15
2.5.	Program Studi .....	16

<b>BAB III TINJAUAN PUSTAKA</b> .....	19
<b>3.1. Praproses</b> .....	19
3.1.1    Pembersihan data.....	19
3.1.2    Case Folding.....	19
3.1.3    Tokenisasi.....	19
3.1.4    Stemming.....	19
3.1.5    Stopword Removing.....	20
<b>3.2. Segmentasi Teks</b> .....	20
<b>3.3. Word Embedding</b> .....	20
<b>3.4. Cosinesimilarit</b> .....	21
<b>3.5. Word Cloud</b> .....	21
<b>BAB IV METODOLOGI</b> .....	23
<b>4.1 Dataset</b> .....	23
<b>4.2 Perancangan Sistem</b> .....	26
4.2.1    Segmentasi tek.....	26
4.2.2    Praproses.....	27
4.2.3    Pemberian Label pada Segmentasi.....	29
4.2.4    Perhitungan kemiripan.....	31
<b>4.3 Implementasi</b> .....	33
4.3.1    Input Data.....	33
4.3.2    Segmentasi Teks.....	37
4.3.3    Praproses.....	42
4.3.4    Wordcloud.....	49

<b>4.3.5</b>	<b>Melabelkan Hasil Segment .....</b>	<b>51</b>
<b>4.3.6</b>	<b>Cosine Similarity.....</b>	<b>57</b>
<b>BAB V</b>	<b>PENGUJIAN DAN EVALUASI .....</b>	<b>61</b>
<b>5.1.</b>	<b>Tujuan Pengujian .....</b>	<b>61</b>
<b>5.2.</b>	<b>Skenario Pengujian .....</b>	<b>61</b>
<b>5.3.</b>	<b>Hasil dan Evaluasi Pengujian .....</b>	<b>61</b>
<b>BAB VI</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>63</b>
<b>6.1.</b>	<b>Kesimpulan .....</b>	<b>63</b>
<b>6.2.</b>	<b>Saran.....</b>	<b>63</b>
<b>DAFTAR PUSTAKA</b>	<b>.....</b>	<b>65</b>
<b>BIODATA PENULIS I</b>	<b>.....</b>	<b>69</b>

*[Halaman ini sengaja dikosongkan]*



## DAFTAR GAMBAR

Gambar 1 Logo ITS.....	6
Gambar 2 flowchart Perancangan Sistem .....	26
Gambar 3 Hasil Segmentasi.....	27
Gambar 4 Hasil Praproses .....	28
Gambar 5 Hasil Pelabelan.....	30
Gambar 6 Hasil cosine similarity .....	32
Gambar 7 gabungan Dataset .....	36
Gambar 8 Membuang Tabel yang tidak diperlukan .....	37
Gambar 9 Hasil segmentasi .....	41
Gambar 10 Memberi Id_kalimat .....	42
Gambar 11 Hasil Membuat Daftar Kata .....	44
Gambar 12 Preprocessing.....	46
Gambar 13 hasil fungsi filter daftar frequency .....	50
Gambar 14 menampilkan daftar frequency.....	51
Gambar 15 Menyatukan data berdasarakan segment .....	53
Gambar 16 Prediksi label berdasarkan segment .....	56
Gambar 17 grafic perhitungan cosine.....	59
Gambar 18 hasil threshold 0.25 keatas .....	61

*[Halaman ini sengaja dikosongkan]*

## DAFTAR TABEL

Table 1 Dataset.....	23
----------------------	----

*[Halaman ini sengaja dikosongkan]*

## DAFTAR KODE SUMBER

Kode sumber 4.1 Import Library .....	34
Kode sumber 4.2 menghubungkan google colab ke google drive .....	34
Kode sumber 4.3 Memuat Data dan menggabungkan data .....	35
Kode sumber 4.4 Membuang Tabel yang tidak diperlukan	37
Kode sumber 4.5 lowercase .....	38
Kode sumber 4.6 Split dataset.....	38
Kode sumber 4.7 Proses Segment teks .....	39
Kode sumber 4.8 Membuat table Segment teks .....	40
Kode sumber 4.9 Memberi topic_set .....	40
Kode sumber 4.10 Memberi Id_kalimat .....	42
Kode sumber 4.11 Memuat Daftar Kata .....	43
Kode sumber 4.12 Preprocessing .....	45
Kode sumber 4.13 Membuat Daftar Kata.....	47
Kode sumber 4.14 Membuat Frequency.....	49
Kode sumber 4.15 Fungsi filter daftar frequency .....	49
Kode sumber 4.16 menampilkan daftar frequency.....	51
Kode sumber 4.17 Menyatukan data berdasarkan segment .....	53
Kode sumber 4.18 Prediksi label berdasarkan segment ...	55
Kode sumber 4.19 Perhitungan cosine .....	58

*[Halaman ini sengaja dikosongkan]*

**LEMBAR PENGESAHAN  
KERJA PRAKTIK**

**Prediksi Teks Wawancara Medis Berdasarkan Nilai  
Kemiripan menggunakan Cosine Similarity**

Oleh:

I Kadek Ricky Suirta

05111840000146

Disetujui oleh Pembimbing Kerja Praktik:

1. Dini Adni Navastara,  
S.Kom., M.Sc.  
NIP. 198510172015042001



(Pembimbing Departemen)

2. Dr. Ir. Diana Purwitasari,  
S.Kom., M.Sc.  
NIP. 197804102003122001



(Pembimbing Lapangan)

**4 Agustus 2023**

*[Halaman ini sengaja dikosongkan]*



## **Prediksi Teks Wawancara Medis Berdasarkan Nilai Kemiripan menggunakan Cosine Similarity**

Nama Mahasiswa : I Kadek Ricky Suirta  
NRP : 05111840000146  
Departemen : Teknik Informatika FTEIC-ITS  
Pembimbing Departemen : Dini Adni Navastara, S.Kom., M.Sc.  
Pembimbing Lapangan : Dr. Ir. Diana Purwitasari, S.Kom., M.Sc.

### **ABSTRAK**

Konsultasi kesehatan *online* semakin banyak digunakan terlebih selama masa pandemi COVID19. Proses tanya jawab yang menjadi bagian dari *interview* dokter dan pasien sebelum dilakukan diagnosis penyakit dimungkinkan tidak terjadi timbal balik pada media *online* tersebut. Masyarakat pada umumnya akan menuliskan lebih dari satu pertanyaan kepada dokter terkait satu kategori penyakit secara luas. Redaksi kalimat pada teks tersebut sering tidak terstruktur dan bercampur antara pertanyaan serta pernyataan. Banyaknya jumlah penanya dengan pertanyaan yang mungkin masih memiliki kemiripan membuat data konsultasi kesehatan terlihat tidak tertata dengan rapi. Pada Tugas Akhir ini akan dilakukan tahapan menggunakan pendekatan segmentasi teks yang dapat membantu pengguna atau pengunjung media konsultasi kesehatan online tersebut dalam mencari data pertanyaan terdahulu.

***Kata Kunci : Konsultasi Kesehatan, Cosine Similarity, Segmentasi Teks***

*[Halaman ini sengaja dikosongkan]*

## **Prediction of Medical Interview Texts Based on Similarity Values using Cosine Similarity**

Student Name : I Kadek Ricky Suirta  
NRP : 05111840000146  
Department : Informatics Engineering FTEIC-ITS  
Department Supervisor : Dini Adni Navastara, S.Kom., M.Sc.  
Field Advisor : Dr. Ir. Diana Purwitasari, S.Kom.,  
M.Sc.M.Sc.

### **ABSTRACT**

Online health consultations are increasingly being used, especially during the COVID19 pandemic. The question and answer process which is part of the doctor and patient interview before a disease diagnosis is made is possible that there will be no reciprocity in the online media. The general public will generally write more than one question to a doctor regarding one broad category of disease. Sentences in the text are often unstructured and mixed up between questions and statements. The large number of questioners with questions that may still have similarities makes the health consultation data look unorganized. This Final Project will carry out stages using a text segmentation approach that can help users or visitors to the online health consultation media in finding previous question data.

***Keywords: Health Consultation, Cosine Similarity, Text Segmentation***

*[Halaman ini sengaja dikosongkan]*

## **KATA PENGANTAR**

Puji syukur penulis panjatkan kepada Tuhan YME atas penyertaan dan karunia-Nya sehingga penulis dapat menyelesaikan salah satu kewajiban penulis sebagai mahasiswa Departemen Teknik Informatika ITS yaitu Kerja Praktik yang berjudul: Prediksi Teks Wawancara Medis Berdasarkan Nilai Kemiripan menggunakan Cosine Similarity

Penulis menyadari bahwa masih banyak kekurangan baik dalam melaksanakan kerja praktik maupun penyusunan buku laporan kerja praktik ini. Namun penulis berharap buku laporan ini dapat menambah wawasan pembaca dan dapat menjadi sumber referensi.

Melalui buku laporan ini penulis juga ingin menyampaikan rasa terima kasih kepada orang-orang yang telah membantu menyusun laporan kerja praktik baik secara langsung maupun tidak langsung antara lain:

1. Kedua orang tua penulis.
2. Ibu Dini Adni Navastara, S.Kom., M.Sc. selaku dosen pembimbing kerja praktik sekaligus koordinator kerja praktik.
3. Ibu r. Ir. Diana Purwitasari, S.Kom., M.Sc. selaku pembimbing lapangan selama kerja praktik berlangsung.

Surabaya, 30 Desember 2022

I Kadek Ricky Suirta

*[Halaman ini sengaja dikosongkan]*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Perkembangan teknologi memberikan kemudahan masyarakat dalam mencari informasi awal misal pada topik kesehatan dari internet. Sebagai contoh, aplikasi *online* konsultasi kesehatan antara lain Halodoc memberikan alternatif bagi masyarakat untuk mencari tahu info terkait masalahnya sebelum melakukan janji temu dengan dokter untuk diagnosis penyakit. Aplikasi Halodoc yang populer di Indonesia memungkinkan pengguna mengajukan pertanyaan pada *platform* tersebut dan akan dijawab oleh dokter-dokter yang bersangkutan dalam bentuk teks.

Selama masa pandemi COVID19, konsultasi kesehatan online semakin banyak digunakan (Purnat, 2020). Proses tanya jawab yang menjadi bagian dari wawancara dokter dan pasien sebelum dilakukan diagnosis penyakit dimungkinkan tidak terjadi timbal balik pada media online tersebut. Masyarakat pada umumnya akan menuliskan lebih dari satu pertanyaan kepada dokter terkait satu kategori penyakit secara luas. Redaksi kalimat pada teks tersebut sering tidak terstruktur dan bercampur antara pertanyaan serta pernyataan. Banyaknya jumlah penanya dengan pertanyaan yang mungkin masih memiliki kemiripan membuat data konsultasi kesehatan terlihat tidak tertata dengan rapi (Chen, Guo, Wu, & Ju, 2020). Data tanya jawab pada media konsultasi online sering ditampilkan berdasarkan kategori yang cakupannya masih cukup luas danurut berdasarkan waktu. Pengunjung enggan untuk melihat data tanya jawab yang sudah ada dan lebih memilih langsung menuliskan teks pertanyaan.

Meski terlihat tidak teratur, namun berdasarkan faktor fungsi wawancara medis, ada beberapa tahapan saat dokter melakukan *interview* dengan pasien: mengembangkan hubungan dengan pasien (*foster relationship*), mengumpulkan informasi (*gathering information*), menyediakan informasi (*providing information*), menyampaikan diagnosis (*decision making*), menentukan terapi

(*enabling treatment related behaviour*), dan memahami respon pasien atas diagnosis (*responding to emotions*) (Ma & Hoppe, 2013). Fungsi wawancara medis tersebut dapat terjadi timbal balik dan tidak selalu beraturan saat konsultasi tidak dilakukan melalui teks tanya jawab. Pada Tugas Akhir ini akan dilakukan tahapan dengan pendekatan segmentasi teks berdasarkan fungsi wawancara medis untuk membantu kategorisasi data tanya jawab.

## **1.2. Tujuan**

Tujuan kerja praktik ini adalah melakukan segmentasi teks berdasarkan faktor fungsi wawancara medis yang digunakan untuk membantu pengkategorian data tanya jawab.

## **1.3. Manfaat**

Manfaat yang diperoleh dengan adanya hasil segmen teks dengan label sesuai fungsi wawancara medis dapat menjadi kamus pertanyaan serta variasi diagnosis dari suatu kategori penyakit yang tertata dari cakupan luas sampai lebih spesifik.

## **1.4. Rumusan Masalah**

Rumusan masalah dari kerja praktik ini adalah sebagai berikut:

1. Bagaimana melakukan pemrosesan data sebelum dilakukan segmentasi teks??
2. Bagaimana melakukan segmentasi teks dan memberikan label per segmen berdasarkan fungsi wawancara medis?
3. Bagaimana melakukan evaluasi hasil Prediksi Teks Wawancara Medis Berdasarkan Nilai Kemiripan menggunakan Cosine Similarity?



## **1.5. Lokasi dan Waktu Kerja Praktik**

Kerja praktik ini dilaksanakan pada waktu dan tempat sebagai berikut:

Lokasi	: Online
Waktu	: 1 Oktober – 14 Mei 2023
Hari Kerja	: Senin – Sabtu
Jam Kerja	: Fleksibel (minimal 12 jam per minggu)

## **1.6. Metodologi Kerja Praktik**

Metodologi dalam pembuatan buku kerja praktik meliputi :

### **1.6.1. Perumusan Masalah**

Untuk mengetahui permasalahan apa yang dapat diselesaikan. Kami melakukan pertemuan Bersama Bu Dini dan Bu Diana untuk merumuskan apa saja yang harus dikerjakan.

### **1.6.2. Studi Literatur**

Setelah mendapat gambaran bagaimana sistem yang akan dibuat, kami melakukan studi literatur mengenai implementasinya. Pada tahap ini dilakukan pencarian, pembelajaran, dan pengumpulan informasi mengenai Segmentasi teks yang mudah dipahami dari internet maupun dari penelitian sebelumnya.

### **1.6.3. Implementasi Sistem**

Implementasi merupakan realisasi dari tahap perancangan. Pada tahap ini kami melakukan eksekusi Segmentasi text pada dataset kami.

### **1.6.4. Pengujian dan Evaluasi**

Setelah hasil dari Segmentasi teks telah didapatkan akan dilakukan pelabelan berdasarkan factor fungsi wawancara medis.

### **1.6.5. Kesimpulan dan Saran**

Pengujian yang dilakukan ini akan dihasilkan kesimpulan dan saran dari pengalaman Kerja Praktik ini.

## **1.7. Sistematika Laporan**

### **1.7.1. Bab I Pendahuluan**

Bab ini berisi latar belakang, tujuan, manfaat, rumusan masalah, lokasi dan waktu kerja praktik, metodologi, dan sistematika laporan.

### **1.7.2. Bab II Profil Perusahaan**

Bab ini berisi gambaran umum Departemen Teknik Informatika Fakultas Teknologi Elektro dan Informatika Cerdas Institut Teknologi Sepuluh Nopember, tempat melakukan kerja praktik.

### **1.7.3. Bab III Tinjauan Pustaka**

Bab ini berisi dasar teori dari teknologi yang digunakan dalam menyelesaikan proyek kerja praktik.

### **1.7.4. Bab IV Metodologi**

Bab ini berisi uraian tahap - tahap yang dilakukan untuk proses implementasi segmentasi teks.

### **1.7.5. Bab V Pengujian dan Evaluasi**

Bab ini berisi hasil uji coba dan evaluasi dari aplikasi yang telah dikembangkan selama pelaksanaan kerja praktik.

### **1.7.6. Bab VI Kesimpulan dan Saran**

Bab ini berisi kesimpulan dan saran yang didapat dari proses pelaksanaan kerja praktik.

## **BAB II**

### **PROFIL PERUSAHAAN**

#### **2.1. Profil Departemen Informatika ITS**

Departemen Teknik Informatika merupakan salah satu departemen besar di Institut Teknologi Sepuluh Nopember (ITS) yang memiliki reputasi sangat baik pada level nasional dan diakui di level internasional. Mulai tahun 2020, Departemen Teknik Informatika merupakan bagian dari Fakultas Teknologi Elektro dan Informatika Cerdas (FT-EIC). Departemen Teknik Informatika ITS memiliki program studi S1, S2, dan S3 yang didukung oleh staf pengajar yang berkualitas yaitu 5 Guru Besar dan 22 dosen dengan kualifikasi S3 dari lulusan universitas-universitas terkemuka di dalam maupun luar negeri. Pada tahun 2019, Departemen Teknik Informatika membuka Kelas Internasional (IUP) didukung program joint degree dengan University of Queensland, University of Newcastle, Fontys University, dan sebagainya (Dr.Eng. Chastine Fatichah, 2020).

Banyak prestasi yang telah dicapai oleh mahasiswa Teknik Informatika ITS baik tingkat nasional maupun internasional seperti Gemastik, ImagineCup Indonesia, serta kompetisi internasional RoboBoat dan RoboCup. Lulusan Teknik Informatika ITS memiliki kompetensi yang unggul dalam bidang informatika ditambah dengan kemampuan interpersonal, teamwork, manajerial, dan technopreneurship yang telah bekerja di perusahaan nasional maupun multinasional dengan rata-rata waktu tunggu mendapatkan pekerjaan kurang dari 2 bulan setelah lulus.

Teknik Informatika memiliki 8 (delapan) bidang minat / rumpun matakuliah yaitu Algoritma dan Pemrograman (AP), Arsitektur dan Jaringan Komputer (AJK), Grafika, Interaksi, dan Game (GIGa), Komputasi Berbasis Jaringan (KBJ), Komputasi Cerdas dan Visi (KCV), Manajemen Cerdas Informasi (MCI), Pemodelan dan Komputasi Terapan (PKT), serta Rekayasa Perangkat Lunak (RPL).



*Gambar 1 Logo ITS*

## **2.2. Lokasi**

Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota Surabaya,  
Jawa Timur 60111

## **2.3. Visi dan Misi**

### **2.3.1 Visi**

Sejalan dengan visi ITS yaitu menjadi perguruan tinggi dengan reputasi internasional dalam ilmu pengetahuan, teknologi, dan seni, terutama yang menunjang industry dan kelautan yang berwawasan lingkungan, maka visi Departemen Informatika adalah menjadi innovator bidang informatika yang unggul di tingkat nasional dengan reputasi nasional dengan reputasi

internasional, serta berperan aktif dalam upaya memajukan dan mensejahterakan bangsa

Visi PSTI adalah menjadi Lembaga Pendidikan dan penelitian di bidang informatika yang unggul di tingkat nasional dan memiliki reputasi internasional (DPTSI, n.d.)

### **2.3.2 Misi**

Departemen Informatika memiliki misi sebagai berikut:

- Menyelenggarakan proses pembelajaran yang berkualitas dan memenuhi standar nasional maupun internasional.
- Melaksanakan penelitian yang inovatif, bermutu, dan bermanfaat.
- Meningkatkan pemanfaatan teknologi informasi dan komunikasi untuk masyarakat
- Menjalin kemitraan dengan berbagai Lembaga, baik di dalam maupun diluar negeri (DPTSI, n.d.).

### **2.3.3 Tujuan**

Tujuan dari PSTI adalah:

- Menghasilkan lulusan kompeten di bidang Informatika, serta memiliki daya saing dan kemandirian untuk berkompetisi di tingkat nasional dan internasional.
- Melakukan perbaikan proses pembelajaran secara berkesinambungan.
- Menghasilkan karya penelitian yang inovatif dan bermanfaat bagi masyarakat, serta peblikasi di jurnal nasional ataupun internasional bereputasi.
- Memberikan kontribusi bagi peningkatan mutu kehidupan masyarakat.

- Mengambil peran aktif dalam kegiatan bidang Informatika pada tingkat nasional dan internasional (DPTSI, n.d.).

### **2.3.4 Sasaran**

- Peningkatan kualitas PBM, penelitian, dan kerjasama.
- Peningkatan kualitas lulusan yang unggul pada tingkat nasional.
- Pertumbuhan dan keberlanjutan (growth and Sustainability), yaitu Pengembangan unggulan dan dikenalnya Jurusan di tingkat internasional.
- Peningkatan citra PSTI sebagai unggulan di tingkat nasional dan semakin dikenalnya PSTI di tingkat internasional (DPTSI, n.d.).

## **2.4. Laboratorium Bidang Minat**

### **2.4.1 Lab. Rekayasa Perangkat Lunak (RPL)**

Di Laboratorium ini ditawarkan bidang minat yang berfokus pada keahlian melakukan pengujian perangkat lunak, Kemampuan mengelola proyek perangkat lunak, Kemampuan mengurangi resiko kesalahan perangkat lunak, dan Kemampuan membuat perangkat lunak game (Dr. Ir. Siti Rochimah MT., n.d.).

Mata kuliah bidang keahlian RPL, antara lain:

- Penjamin Mutu Perangkat Lunak,
- Arsitektur Perangkat Lunak,
- Evolusi Perangkat Lunak,
- Konstruksi Perangkat Lunak
- Penyempurnaan Proses Perangkat Lunak,
- Ekonomi Rekayasa Perangkat Lunak,
- Topik Khusus RPL.

Fasilitas

Laboratorium Rekayasa Perangkat Lunak dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, sampai Intel® Xeon® E5-2640 dengan RAM 4GB-16GB.
- Untuk HDD sebagian besar minimal 1TB.
- Semua monitor berukuran 19" untuk memudahkan mahasiswa dalam melakukan penelitian dan pembelajaran rekayasa perangkat lunak ataupun pemrograman.
- Dilengkapi LED TV 55" untuk mahasiswa dalam melakukan demo pembelajaran.

#### **2.4.2 Lab. Komputasi Berbasis Jaringan (KBJ)**

Di Laboratorium ini ditawarkan bidang keahlian yang ditekankan pada Kemampuan lulusan sarjana/magister/doktor dalam membangun infrastruktur jaringan yang aman, kemampuan membangun sistem grid, Kemampuan membangun aplikasi jaringan sesuai Standard dan Kemampuan membangun aplikasi multimedia berbasis jaringan (Prof. Tohari Ahmad, n.d.).

Mata kuliah bidang keahlian KBJ, antara lain:

- Komputasi Bergerak.
- Sistem Terdistribusi.
- Keamanan Informasi dan Jaringan.
- Jaringan Multimedia.
- Komputasi Grid dan Paralel.
- Kompresi Data.
- Komputasi Awan.
- Forensik Digital.
- Komputasi Pervasif dan Jaringan Sensor.
- Topik Khusus Komputasi berbasis Jaringan.

## Fasilitas

Laboratorium Komputasi Berbasis Jaringan dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, sampai Intel® Xeon® E5-2640 dengan RAM 4GB-16GB.
- Untuk HDD sebagian besar minimal 1TB.
- Semua monitor berukuran 19" untuk memudahkan mahasiswa dalam melakukan penelitian dan pembelajaran rekayasa perangkat lunak ataupun pemrograman.
- Dilengkapi LED TV 55" untuk mahasiswa dalam melakukan demo pembelajaran.

### **2.4.3 Lab. Komputasi Cerdas dan Visi (KCV)**

Di Laboratorium ini ditawarkan bidang keahlian yang ditekankan pada kemampuan lulusan dalam memanipulasi dan menganalisis data citra pada berbagai bidang aplikasi (a.l. biomedika, industri), kemampuan menerapkan metode sistem cerdas pada berbagai bidang aplikasi dan kemampuan memodelkan dan mengoptimasikan sistem nyata (Prof. Dr.Eng. Nanik Suciati, n.d.).

Mata kuliah bidang keahlian KCV, antara lain:

- Pengolahan Citra Digital,
- Analisis Data Multivariat,
- Data Mining,
- Komputasi Biomedik,
- Visi Komputer,
- Sistem Temu Kembali Informasi,
- Robotika,
- Analisis Media Sosial dan Topik Khusus KCV.

## Fasilitas



Laboratorium Komputasi Cerdas dan Visi dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, Intel® Xeon® E5-2640 dengan RAM 4GB-16GB, Sampai dengan Processor i9 Generasi 12 GPU 3080TI
- Untuk HDD sebagian besar minimal 1TB.
- Semua monitor berukuran 19" untuk memudahkan mahasiswa dalam melakukan penelitian dan pembelajaran.
- Dilengkapi LED TV 55" untuk mahasiswa dalam melakukan demo pembelajaran.
- LCD Projector untuk memudahkan mahasiswa yang membutuhkan beberapa display.
- Ploter Ungut menunjang kebutuhan mahasiswa mencetak hasil tugas yang dibutuhkan.

#### **2.4.4 Lab. Arsitektur dan Jaringan Komputer (AJK)**

Laboratorium di bidang minat ini menawarkan bidang keahlian yang ditekankan pada Kemampuan lulusan dalam membangun berbagai macam arsitektur jaringan sesuai standar teknologi terkini dan menerapkan keamanan jaringan (Dr. Wahyu Suadi, n.d.).

Mata kuliah bidang keahlian AJK, antara lain:

- Sistem Operasi.
- Jaringan Komputer.
- Pemrograman Jaringan.
- Jaringan Nirkabel.
- Teknologi Antar Jaringan.
- Perancangan Keamanan Sistem Dan Jaringan.
- Desain dan Audit Jaringan.
- Teknologi Internet of Things.
- Topik Khusus Arsitektur Dan Jaringan Komputer.

Fasilitas

Laboratorium Arsitektur dan Jaringan Komputer memiliki layout seperti ruang kelas, sehingga laboratorium ini sering digunakan untuk perkuliahan, praktikum, workshop, Computer Based Test (CBT) dll.

- Computer Client i3 – i5 dengan ram minimal 4 GB.
- 2 Unit Server Praktikum, Xeon E3 1220 dengan ram 16 GB.
- 1 Unit Server GPU Prosesor i9.
- Memiliki ruang server yang standby 24 jam.
- Memiliki Ruang Rapat / Ruang Sidang dan Smart TV.
- Dilengkapi dengan Audio System
- dan LCD Projector.

#### **2.4.5 Lab. Grafika Interaksi dan Game (GIGa)**

Laboratorium ini di bidang minat ini menawarkan bidang keahlian yang ditekankan pada kemampuan lulusan dalam mendesain, mengembangkan dan mendokumentasikan proses pembuatan game sesuai dengan standar. Serta membuat model 3 dimensi dan pemograman di dalam realitas virtual serta aplikasi realitas virtual 3 dimensi dengan menggunakan game engine (Imam Kuswardayan, n.d.).

Mata kuliah bidang keahlian GIGa antara lain:

- Teknik Pengembangan Game,
- Sistem Game,
- Animasi Komputer dan Pemodelan 3D,
- Realitas Virtual dan Augmentasi, dan
- Topik Khusus GIGa.

Fasilitas

- HP PC 690-0017D + MONITOR 21” ( HP Pavilion Gaming Desktop)
- Logitech Webcam B525

- Oculus Rift Set (Game controller)
- Oculus VR Development KIT DK-2-V51-BDL (Game controller)
- HTC VIVE Virtual Reality (Game controller)
- Logitech Gaming Wheel
- PC High End (Intel Core i9-12900K, 64GB DDR5, RTX 3080Ti 12GB)

#### **2.4.6 Lab. Algoritma dan Pemrograman (AP)**

Laboratorium ini menawarkan bidang keahlian yang ditekankan pada kemampuan lulusan dalam (Dr. Yudhi Purwananto, n.d.),

- untuk merancang dan menganalisa algoritma dalam menyelesaikan permasalahan secara efektif dan efisien berdasarkan kaidah pemrograman yang kuat,
- mampu mengaplikasikan model pemrograman yang mendasari berbagai bahasa pemrograman yang ada,
- serta mampu memilih bahasa pemrograman untuk menghasilkan aplikasi yang sesuai, seperti mengembangkan sistem/aplikasi berbasis kerangka kerja dan pada perangkat bergerak.

#### **Mata kuliah wajib terkait AP antara lain:**

Dasar Pemrograman, Struktur Data, Pemrograman Berorientasi Objek, Perancangan dan Analisis Algoritma, Pemrograman Web, Pemrograman Berbasis Kerangka Kerja

#### **Mata kuliah bidang keahlian AP antara lain:**

Pengembangan Analisis Algoritma, Pemrograman Berbasis Antarmuka, Pemrograman Perangkat Bergerak, Topik Khusus Algoritma dan Pemrograman.

Fasilitas

Laboratorium Algoritma dan Pemrograman dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, sampai Intel® Xeon® E5-2640 dengan RAM 4GB-16GB.
- Untuk HDD sebagian besar minimal 1TB.
- Semua monitor berukuran 19" untuk memudahkan mahasiswa dalam melakukan kegiatan pemrograman.
- PC High End (Intel Core i9-12900K, 64GB DDR5, RTX 3080Ti 12GB)

#### **2.4.7 Lab. Manajemen Cerdas Informasi (MCI)**

Laboratorium di bidang minat ini menawarkan bidang keahlian yang ditekankan pada kemampuan lulusan dalam menganalisis, mensintesa dan mengevaluasi proses bisnis dan sistem informasi pada sistem Enterprise, mengimplementasikan rekayasa pengetahuan ke dalam suatu aplikasi, melakukan investigasi, pengujian, evaluasi kematangan dan kepatutan terhadap prosedur standard dan tata kelola teknologi informasi, melakukan tata kelola proyek dan sumber daya manusia dan merancang dan mengimplementasikan solusi basis data terdistribusi dan teknologi Big Data (Prof. Drs. Ec. Ir. Riyanarto Sarno, n.d.).

Mata kuliah bidang keahlian MCI, antara lain:

- Sistem Enterprise,
- Rekayasa Pengetahuan,
- Sistem Informasi Geografis,
- Audit Sistem,
- Tata Kelola Teknologi Informasi,
- Basis Data Terdistribusi,
- Big Data,
- Topik Khusus Manajemen Informasi.

## Fasilitas

Laboratorium Manajemen Informasi Cerdas dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, sampai Intel® Xeon® E5-2640 dengan RAM 4GB-16GB.
- Untuk HDD sebagian besar minimal 1TB.
- LCD Proyektor
- Android TV 43"
- Android TV 65"
- Laptop High End

### **2.4.8 Lab. Pemodelan dan Komputasi Terapan (PKT)**

Laboratorium ini memfasilitasi riset dan kerjasama industri di bidang pemodelan & simulasi, peramalan sains, optimasi, serta komputasional saintifik.

Beberapa mata kuliah pendukung yg diajarkan antara lain Matematika Diskrit, Aljabar Linier & Matriks, Teori Graph & Otomata, Komputasi Numerik, Probabilistik & Statistik, Riset Operasi, Pemodelan & Simulasi, Analisis Data Multivariat. Serta beberapa mata kuliah pasca sarjana lainnya (Victor Hariadi, n.d.).

bidang keahlian & riset

- Prof.Dr.Ir. Joko Lianto Buliali (pemodelan & simulasi)
- Dr. Ahmad Saikhu, S.Si., MT. (peramalan saintifik)
- Bilqis Amaliah, S.Kom.,M.Kom. (teknik optimasi)
- Victor Hariadi, S.Si., M.Kom. (pemodelan otomata)
- Yudhi Purwananto, S.Kom., M.Kom. (komputasi saintifik)
- Arya Yudhi Wijaya, S.Kom, M.Kom. (pemodelan graph)

## Fasilitas

Laboratorium Pemodelan dan Komputasi Terapan dilengkapi dengan komputer yang memiliki spesifikasi bervariasi mulai dari,

- Processor Intel Core i3 Gen-3, i5 Gen-8, sampai Intel® Xeon® E5-2640 dengan RAM 4GB-16GB.
- Untuk HDD sebagian besar minimal 1TB.
- Semua monitor berukuran 19" untuk memudahkan mahasiswa dalam melakukan penelitian dan pembelajaran.
- Dilengkapi LED TV 55" untuk mahasiswa dalam melakukan demo pembelajaran.
- LCD Projector untuk memudahkan mahasiswa yang membutuhkan beberapa display.
- Plotter Untuk menunjang kebutuhan mahasiswa mencetak hasil tugas yang dibutuhkan.

## **2.5. Program Studi**

Informatika adalah disiplin ilmu pengetahuan yang melibatkan praktik pemrosesan informasi dan rekayasa sistem. Disiplin ilmu ini merupakan terapan dari sains informasi yang mempertimbangkan interaksi antara manusia dan informasi pada konstruksi antar muka, teknologi, dan sistem. Seiring dengan kemajuan teknologi komputer, kuantitas dan kualitas pemrosesan informasi secara digital oleh individual dan organisasi turut meningkat. Hal ini menuntut pembelajaran dan penelitian pada disiplin ilmu Informatika dari segi komputasi, matematika, biologi, kognitif, dan aspek sosial, termasuk studi terhadap dampak sosial dari teknologi informasi.

Di Departemen Teknik Informatika, selain meningkatkan ilmu perihal Informatika, mahasiswa juga belajar untuk menjadi

kreatif dan inovatif, memiliki sifat kepemimpinan, terintegrasi, serta dapat bekerja dengan baik dalam sebuah tim. Departemen Teknik Informatika memiliki misi untuk memberikan kualitas pendidikan yang sangat baik dalam hal Informatika, mengembangkan dan menerapkan Informatika yang dapat bermanfaat untuk masyarakat dan negara, serta memberikan kepuasan dan kesejahteraan kepada sivitas Informatika.

*[Halaman ini sengaja dikosongkan]*



## **BAB III**

### **TINJAUAN PUSTAKA**

#### **3.1. Praproses**

Praproses data berguna untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur (Efrizoni, Defit, Tajuddin, & Anggrawan, 2020).

##### 3.1.1 Pembersihan data

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan (Fauziningrum, M.Pd & Sulistyaningsih, 2021). Data berkualitas buruk akan memberikan hasil dan algoritma yang tidak bisa dijamin kebenarannya meski proses analisisnya benar.

##### 3.1.2 Case Folding

Case folding merupakan pengubah semua teks menjadi huruf kecil (Andana, 2021). Misalnya data teks yang berupa tulisan "Makan Nasi" maka dengan case folding artinya mengubah semua huruf menjadi huruf kecil (lowercase) semua.

##### 3.1.3 Tokenisasi

Tokenisasi adalah langkah praproses yang bertujuan agar tahap pemotongan input string berdasarkan setiap kata yang menyusunnya (Honakan et al., 2018). Tokenisasi berguna ketika ingin mengekstrak makna dari sebuah teks. Contohnya dalam sebuah kalimat ingin mendeteksi kata benda dan kata kerja yang ada dalam kalimat tersebut, atau ingin mencari nama orang yang dimation dalam sebuah kalimat. Maka dengan melakukan pemecahan satu per satu, sistem dapat melakukan pengecekan satu per satu juga terhadap tiap-tiap teks yang ada pada kalimat.

##### 3.1.4 Stemming

Stemming berfungsi untuk mengurangi ukuran kosakata dengan memetakan varian berdasarkan kata dasar (Rianto et al., 2021). Sebagai contoh, kata bersama, kebersamaan, menyamai,

akan distem ke root wordnya yaitu "sama". Proses stemming pada teks Bahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan (Guterres et al., 2019)

#### 3.1.5 Stopword Removing

*Stopwords removal* yaitu menghilangkan kata-kata yang sering muncul dan tidak memiliki arti. Contoh stopword adalah menghapus angka, dan menghapusnya dapat mengurangi kebisingan dan meningkatkan efisiensi. Menghilangkan tanda baca, Tanda baca adalah karakter unik seperti tanda seru, koma, tanda tanya, dan lain-lain. Itu tidak diperlukan dalam klasifikasi sentimen. (Pradana & Hayaty, 2019).

### 3.2. Segmentasi Teks

*Text Segmentation* merupakan poses yang digunakan untuk memecah teks menjadi beberapa bagian yang berkelanjutan dan berkoherensi. Langkah-langkah yang akan dilakukan berupa memisahkan text menjadi beberapa  $N$  elemen dasar, kemudian kita memberikan skor untuk setiap kadidat elemen. Dan terakhir kita bisa menentukan untuk memisahkan teks menjadi beberapa bagian. Dalam garis besar *text segmentation* dibutuhkan beberapa langkah, yaitu *representation*, *scoring*, dan *splitting* (Aleml & Ginsparg, 2015).

### 3.3. Word Embedding

*Word Embedding* merupakan salah satu konsep terpenting dalam *NLP* modern (Hagiwara, 2021). *Word embedding* adalah representasi vector terus menerus dari sebuah teks. Setiap kata adalah vector yang merepresentasikan sebuah titik pada space

dengan dimensi tertentu. *Word Embedding* sendiri digunakan untuk representasi makna yang lebih baik karena adanya keterbatasan informasi pada teks pendek (Wang & Liu, 2018)

**3.4. Cosinesimilarit**

Dalam *Text Mining*, *Similarity* (atau *distance*) adalah cara klasik untuk menghitung kesamaan antara dua dokumen teks, dan secara luas digunakan dalam berbagai metode *Machine Learning* (ML), termasuk engelompokan dan klasifikasi. Metode ML membantu dari koleksi yang sangat besar yang dikenal sebaga *Big Data* (Oghbaie & Zanjireh, 2018). yang digunakan untuk menghitung *cosine similarity*.

$$Cosine\ Similarity = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \dots\dots\dots(1)$$

Keterangan:

*Cosine Similarity* : tingkat kemiripan data dengan vector tertentu,

*A* : bobot vektor A,

*B* : bobot vektor B,

*n* : jumlah vektor yang unik dalam data set.

(Saputra & Muttaqin, 2013)

**3.5. Word Cloud**

Word cloud adalah visualisasi yang menampilkan kata-kata dari teks yang diberikan, dengan ukuran font yang lebih besar untuk kata-kata yang lebih sering muncul dalam teks tersebut. Word cloudsalah satu Teknik visualisasi yang menunjukkan hubungan antara jumlah kata dengan memperbesar kata yang lebih sering digunakan dengan cepat (Yanuarti, 2021).

*[Halaman ini sengaja dikosongkan]*

## BAB IV METODOLOGI

Bab ini akan menjelaskan tahap implementasi yang dilakukan.

### 4.1 Dataset

Data yang digunakan untuk penelitian ini diambil dari data tanya jawab Halodoc. Jumlah data Halodoc berjumlah 18.737 data dan sudah dikumpulkan dari tahun 2014-2021. Data yang akan diolah berjumlah 498 dalam format csv. Data tersebut memiliki 5 atribut yaitu "id", "title", "question", "answer", dan "topic\_set", namun data yang diproses adalah atribut "id", "answer" dan "topic\_set"

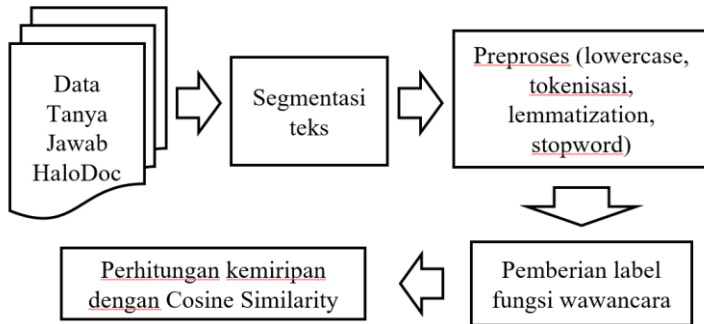
*Table 1 Dataset*

index	answer	Topic_set
0	Halo yayan, Terima kasih atas pertanyaannya. Berdasarkan riwayat keluhan istri Anda yang sudah Anda sampaikan, yaitu muntah - muntah dan diare lebih dari 5-10x, terdapat kemungkinan kondisi yang dialami oleh istri Anda, salah satunya adalah keracunan makanan, infeksi saluran cerna, adanya malapsorbsi / intoleransi makanan yang ditandai dengan gejala - gejala yang Anda sebutkan tersebut. Gejala muntah dan diare pada dasarnya merupakan respon tubuh yang wajar sebagaimana tubuh berusaha mengeluarkan toksin / racun dari makanan / minuman yang terkontaminasi yang dikonsumsi sebelumnya. Kondisi ini dapat membaik dengan sendirinya dalam hitungan beberapa hari tanpa perlu pengobatan khusus dari dokter. Pantau terus gejala yang muncul, kembalikan cairan tubuh yang keluar, dan monitor tanda - tanda dehidrasi seperti	Gangguan Pencernaan

	<p>lemas, mata cowong, tidak buang air kecil lebih dari 6 jam atau urin berwarna pekat, mulut kering dan rasa sangat haus, dan adanya tanda - tanda kebingungan. Selain itu, berikut anjuran yang dapat dilakukan untuk mengurangi keluhan: Menghindari makan terlebih dahulu selama beberapa jam. Minum air, sup, atau minuman elektrolit untuk membantu mengembalikan cairan yang keluar dari tubuh istri Anda. Ketika makan nanti, makan sedikit - sedikit dengan makanan yang tawar seperti roti - rotian, nasi / bubur, biskuit. Istirahat yang cukup. Menghindari susu dan produk susu, kopi, minuman bersoda, dan makanan asam / pedas / berlemak terlebih dahulu. Apabila gejala tidak membaik dalam waktu 3-5 hari, masih terus muntah atau diare dan terdapat darah di dalamnya, demam, lemas, langsung hubungi petugas fasilitas layanan kesehatan terdekat. Semoga informasi ini dapat membantu dan semoga lekas sembuh. Salam.</p>	
1	<p>Halo. Terima kasih atas pertanyaan yang disampaikan ke Alodokter. Kami dapat memahami kekhawatiran yang Anda rasakan. Siklus menstruasi diawali oleh penguasaan sel telur dari indung telur ke rahim, yang diikuti oleh penebalan dinding rahim untuk persiapan menempelnya hasil fertilisasi / pembuahan sel telur oleh sperma. Bila sel telur tidak dibuahi oleh sperma, dinding rahim akan meluruh sehingga keluar melalui vagina dalam bentuk darah serta sel - sel yang rusak. Siklus menstruasi biasanya berlangsung setiap 21-35 hari, dengan haid yang berlangsung selama 2-8 hari, dan volume darah haid berkisar dari 20-60 mL (setara dengan 1-6 pembalut sehari). Tidak menutup kemungkinan bahwa gejala ketidakteraturan menstruasi yang Anda alami merupakan bentuk gangguan haid. Ada beberapa kemungkinan bentuk gangguan haid, yang tersering antara lain: Metroragia, yaitu perdarahan</p>	Intim Wanita

	<p>haid yang interval antarhaidnya tak teratur. Menometroragia, yaitu peningkatan volume / durasi perdarahan haid yang disertai interval antarhaid yang tidak teratur. Hipermenorea/ menoragia, yaitu peningkatan volume perdarahan haid (lebih dari 80 mL darah / hari) dan / atau memanjangnya durasi haid (lebih dari 7 hari), tetapi dengan interval antarhaid yang tetap normal. Ada banyak kemungkinan penyebab gangguan haid, meliputi: Masalah pada organ reproduksi (rahim, leher rahim, saluran / tuba falopi, indung telur), termasuk pula sindrom polikistik ovarium. Faktor - faktor lain di luar sistem organ reproduksi, meliputi: gangguan hormonal, efek samping obat, gangguan pembekuan darah, stres, episode depresi). Supaya kondisi Anda dapat dievaluasi dan ditangani secara tepat, kami menyarankan Anda untuk memeriksakan diri secara langsung ke dokter spesialis kebidanan dan kandungan. Selain anamnesis dan pemeriksaan fisis, dokter mungkin menyarankan pemeriksaan penunjang bila ditemukan indikasi, misalnya: ultrasonografi (USG) kandungan, pemeriksaan darah, atau CT - scan. Untuk membantu redakan keluhan, Anda disarankan untuk melakukan langkah - langkah sederhana berikut ini: Istirahat yang cukup. Jaga berat badan dalam rentang normal, Makan makanan bergizi seimbang secara teratur. Rutin melakukan latihan fisik atau olahraga. Minum air putih paling tidak 8 gelas sehari Manajemen stres dengan baik. Untuk memperkaya wawasan, Anda dapat membaca artikel tentang gangguan menstruasi. Demikian informasi dari kami. Mudah - mudahan senantiasa sehat. Semoga bermanfaat.</p>	
--	---	--

## 4.2 Perancangan Sistem



Gambar 2 flowchart Perancangan Sistem

### 4.2.1 Segmentasi teks

Pada proses segmentasi teks, dataset akan dibagi menjadi beberapa bagian yang berkoherensi. Beberapa bagian tersebut akan diberikan skor untuk setiap kandidat elemen sehingga teks dapat dipisahkan menjadi beberapa bagian (Alemi & Ginsparg, Text Segmentation based on Semantic Word Embeddings, 2015)



No_ID	Topic_set	Answer	segment	id_kalimat	
0	0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya	0	1
1	0	ganggu cerna	berdasarkan riwayat keluhan istri anda yang s...	1	2
2	0	ganggu cerna	gejala muntah dan diare pada dasarnya merupak...	1	3
3	0	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	1	4
4	0	ganggu cerna	pantau terus gejala yang muncul, kembalikan c...	2	5
...	...	...	...	...	...
8576	499	batu ginjal	hindari memijat / mengurut bagian perut yang ...	2	9
8577	499	batu ginjal	jika nyeri terasa mengganggu dapat mengonsums...	2	10
8578	499	batu ginjal	baca laman artikel berikut ini: nyeri saat bu...	2	11
8579	499	batu ginjal	demikian semoga bermanfaat	3	12
8580	499	batu ginjal	salam, dr muliani sukiman	3	13

8584 gambar 3.5 selanjutnya

*Gambar 3 Hasil Segmentasi*

## 4.2.2 Praproses

Langkah-langkah standar praproses teks telah diterapkan seperti pembersihan teks, normalisasi teks, mengubah teks menjadi huruf kecil, lemmatization untuk mengubahnya menjadi bentuk dasar dari sebuah kata, dan menghilangkan stopwords karena memiliki pengaruh paling kecil dalam kalimat. Sebelum praproses, dilakukan tokenisasi untuk membagi kalimat menjadi beberapa bagian yang disebut token atau kata-kata dengan fungsi `word_tokenize` dari library Python `nlTK.tokenize`.

No_ID	Topic_set	Answer	ans_clean	segment
0	0	ganggu cerna halo yayan, terima kasih atas pertanyaannya b...	halo yayan terima kasih dasar riwayat keluh is...	0
1	0	ganggu cerna kondisi ini dapat membaik dengan sendirinya d...	kondisi hitung obat khusus dokter pantau gejala...	1
2	0	ganggu cerna selain itu, berikut anjuran yang dapat dilaku...	anjur laku keluh hindar makan jam	2
3	0	ganggu cerna minum air, sup, atau minuman elektrolit untuk...	minum air sup minum elektrolit bantu cair tubu...	3
4	0	ganggu cerna menghindari susu dan produk susu, kopi, minum...	hindar susu produk susu kopi minum soda makan ...	4
...	...	...	...	...
2982	499	batu ginjal hai nyeri perut setelah buang air kecil dapat...	hai nyeri perut buang air infeksi salur kemih ...	0
2983	499	batu ginjal dengan demikian penanganan yang tepat dapat d...	tangan sesuai kondisi dasar anjur istirahat ko...	1
2984	499	batu ginjal perbanyak konsumsi air putih 2-3 liter sehari...	konsumsi air putih liter tahan buang air buang...	2
2985	499	batu ginjal hindari memijat / mengurut bagian perut yang ...	hindar pijat urut perut sakit	3
2986	499	batu ginjal jika nyeri terasa mengganggu dapat mengonsums...	nyeri ganggu konsumsi paracetamol ibuprofen re...	4

2987 rows x 5 columns

*Gambar 4 Hasil Praproses*

### **4.2.3 Pemberian Label pada Segmentasi**

Pada tahap ini dilakukan pemberian label pada Segmentasi berdasarkan daftar frequency yang didapat. Data frequency didapat dari pecahan kata yang sudah dipilah-pilah dari hasil Wawancara Medis.

No_ID	Topic_set	Answer	ans_clean	segment	F1	F2	F3	F4	F5	F6	predicted_label	
0	0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya b...	halo yayan terima kasih dasar riwayat keluh is...	0	4	9	33	32	30	1	F3
1	0	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	kondisi hitung obat khusus dokter pantau gejala...	1	0	9	21	16	28	1	F5
2	0	ganggu cerna	selain itu, berikut anjuran yang dapat dilaku...	anjur laku keluh hindar makan jam	2	0	1	5	4	6	1	F5
3	0	ganggu cerna	minum air, sup, atau minuman elektrolit untuk...	minum air sup minum elektrolit bantu cair tubu...	3	1	3	12	9	19	0	F5
4	0	ganggu cerna	menghindari susu dan produk susu, kopi, minum...	hindar susu produk susu kopi minum soda makan ...	4	8	3	21	15	25	1	F5
...	...	...	...	...	...	...	...	...	...	...	...	...
2982	499	batu ginjal	hai nyeri perut setelah buang air kecil dapat...	hai nyeri perut buang air infeksi salur kemih ...	0	7	16	33	28	32	12	F3
2983	499	batu ginjal	dengan demikian penanganan yang tepat dapat d...	tangan sesuai kondisi dasar anjur istirahat ko...	1	0	2	13	9	13	0	F3
2984	499	batu ginjal	perbanyak konsumsi air putih 2-3 liter sehari...	konsumsi air putih liter tahan buang air buang...	2	0	7	10	9	10	5	F3
2985	499	batu ginjal	hindari memijat / mengurut bagian perut yang ...	hindar pijat urut perut sakit	3	2	1	5	4	5	0	F3
2986	499	batu ginjal	jika nyeri terasa mengganggu dapat mengonsums...	nyeri ganggu konsumsi paracetamol ibuprofen re...	4	9	7	16	11	18	8	F5

Gambar 5 Hasil Pelabelan

#### 4.2.4 Perhitungan kemiripan dengan Cosine Similarity

Pada tahap ini akan dilakukan cara perhitungan kemiripan dengan menggunakan *cosine similarity*, data yang telah disegmentasi akan dilakukan vektorisasi teks terhadap data. Vektorisasi teks tersebut menggunakan *Word2vec* dengan menerapkan model *skip-gram* dan model *CBOw (Continuous Bag-of-Words)* untuk menghasilkan ruang vektor yang dapat berjumlah beberapa ratus dimensi (Rahutomo, Ikawati, & Rohman, 2019). Kemudian akan dilakukan perhitungan *cosine similarity* untuk menghitung kemiripan antara kalimat satu dan kalimat yang lain. Persamaan 1. yang digunakan untuk menghitung *cosine similarity*.

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \dots\dots\dots(1)$$

Keterangan:

- Cosine Similarity* : tingkat kemiripan data dengan vector tertentu,
- A* : bobot vektor A,
- B* : bobot vektor B,
- n* : jumlah vektor yang unik dalam data set. (Saputra & Muttaqin, 2013)

No_ID	Topic_set	Answer	ans_clean	segment	F1	F2	F3	F4	F5	F6	predicted_label	cosine_similarity	
0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya b...	halo yayan terima kasih dasar riwayat keluh is...	0	4	9	33	32	30	1	F3	0.25855	
1	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	kondisi hitung obat khusus dokter pantau gejala...	1	0	9	21	16	28	1	F5	0.208075	
2	ganggu cerna	selain itu, berikut anjuran yang dapat dilaku...	anjur laku keluh hindar makan jam	2	0	1	5	4	6	1	F5	0.361508	
3	ganggu cerna	minum air, sup, atau minuman elektrolit untuk...	minum air sup minum elektrolit bantu cair tubu...	3	1	3	12	9	19	0	F5	0.21174	
4	ganggu cerna	menghindari susu dan produk susu, kopi, minum...	hindar susu produk susu kopi minum soda makan ...	4	8	3	21	15	25	1	F5	0.176445	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2982	499	batu ginjal	hai nyeri perut setelah buang air kecil dapat...	hai nyeri perut buang air infeksi salur kemih ...	0	7	16	33	28	32	12	F3	0.617022
2983	499	batu ginjal	dengan demikian penanganan yang tepat dapat d...	tangan sesuai kondisi dasar anjur istirahat ko...	1	0	2	13	9	13	0	F3	0.14095
2984	499	batu ginjal	perbanyak konsumsi air putih 2-3 liter sehari...	konsumsi air putih liter tahan buang air buang...	2	0	7	10	9	10	5	F3	0.196326
2985	499	batu ginjal	hindari memijat / mengurut bagian perut yang ...	hindar pijat urut perut sakit	3	2	1	5	4	5	0	F3	0.156676
2986	499	batu ginjal	jika nyeri terasa mengganggu dapat mengonsumsi...	nyeri ganggu konsumsi paracetamol ibuprofen re...	4	9	7	16	11	18	8	F5	0.23453

2987 rows × 13 columns

*Gambar 6 Hasil cosine similarity*

## 4.3 Implementasi

Kali ini akan dijelaskan langkah-langkah dalam mengerjakan *project* ini.

### 4.3.1 Input Data

Pada tahap ini dilakukan penginputan data kedalam googl colab.

- **Import Library**

Langkah pertama pada kode sumber 4.1 dilakukannya *import* library yang dibutuhkan, agar dapat menghasilkan output yang diharapkan ketika program dijalankan.

```
1. import pandas as pd
2. import numpy as np
3. import math
4. import seaborn as sns
5. import nltk
6. import re
7. import csv
8. from tqdm import tqdm
9. import warnings
10. warnings.filterwarnings("ignore")
11. from collections import Counter
12. from nltk.corpus import stopwords
13. from nltk.tokenize import word_tokenize
14. from nltk.tokenize.treebank import
    TreebankWordDetokenizer
15. from nlp_id.lemmatizer import Lemmatizer
```

```

16. from sklearn.metrics.pairwise import
    cosine_similarity
17. from sklearn.feature_extraction.text
    import CountVectorizer
18. import matplotlib.pyplot as plt

19. import os
20. from gensim.models import word2vec
21. from sklearn.feature_extraction.text import
    CountVectorizer
22. %matplotlib inline
23. from textsplint.tools import get_penalty,
    get_segments
24. from textsplint.algorithm import
    split_optimal, split_greedy, get_total
25. import logging
26. from wordcloud import WordCloud
27. from wordcloud import STOPWORDS
28. from wordcloud import ImageColorGenerator
29. import matplotlib.pyplot as plt

```

*Kode sumber 4.1 Import Library*

- **Menyambungkan ke google Drive**

Setelah melakukan import library pada google colab, dilakukannya penyambungan google colab pada google Drive pada kode sumber 4.2.

```

1. from google.colab import drive
2. drive.mount('/content/drive')

```

*Kode sumber 4.2 menghubungkan google colab ke google drive*



- **Memuat Data dari google drive**

Setelah tersambung ke google drive, maka dapat memuat data dengan kode sumber 4.3 dari baris 1-2, yaitu file “DataPertama.xlsx dan DataKedua.xlsx”. Kemudian kedua data akan digabungkan dengan baris 3-6.

```
1. data1 =  
   pd.read_excel('/content/drive/MyDrive/TA &  
   KP/Dataset/249DataPertama.xlsx')  
2. data2 =  
   pd.read_excel('/content/drive/MyDrive/TA &  
   KP/Dataset/249DataKedua.xlsx')  
  
3. data2 ['no'] = data2['no'] + 250  
4. data = [data1, data2]  
5. data = pd.concat(data)  
6. data = data.rename(columns={'no': 'No_Id'})  
7. data
```

*Kode sumber 4.3 Memuat Data dan menggabungkan data*

No_Id	title	question	answer	topic_set	answer_labeled	labeled	
0	0	Mual disertai pusing dan mencret setelah makan...	Halo dok selamat malam, kemarin jam 18.15 saya...	Halo yayan, Terima kasih atas pertanyaannya. B...	Gangguan Pencernaan	[F1]. Halo yayan, Terima kasih atas pertanyaan...	NaN
1	1	Menstruasi tidak teratur berdampak pada psikis	Dok, saya mahasiswa usia 20 tahun. Beberapa bu...	Halo. Terima kasih atas pertanyaan yang disamp...	Intim Wanita	[F1]. Halo. Terima kasih atas pertanyaan yang ...	NaN
2	2	Obat untuk penderita rematik.	Dok, apakah penderita RA harus minum MTX disam...	Selamat pagi, terima kasih atas pertanyaannya...	Obat	[F1]. Selamat pagi, terima kasih atas pertanya...	NaN
3	3	Frekuensi normal BAB dalam sehari	Dok bab saya sehari bisa 3x tapi sedikit kadan...	Halo Friscilla, terima kasih ya sudah bertanya...	Gangguan Pencernaan	[F1]. Halo Friscilla, terima kasih ya sudah be...	NaN
4	4	Rasa panas di punggung saat mengalami asam lam...	alodok, apa iya lambung bermasalah menimbuka...	Halo, terima kasih atas pertanyaannya untuk Al...	Asam Lambung	[F1]. Halo, terima kasih atas pertanyaannya un...	NaN
...	...	...	...	...	...	...	...
245	495	sakit pinggang sebelah kiri disertai kencing t...	Ass...dok mau tx knp pingg seblh kiri skt bt d...	Halo Anatasya..... Terima kasih atas pertany...	batu-ginjal	[F1]. Halo Anatasya..... Terima kasih atas p...	NaN
246	496	sakit saat buang air kecil	Selamat malam dok :) saya Nia dari kalsel :) s...	Hai. Keluhan tersebut dapat disebabkan oleh: i...	batu-ginjal	[F1]. Hai. [F3]. Keluhan tersebut dapat diseba...	NaN
247	497	Sakit pinggang bawah sebelah kanan	Siang dok saya yeni baru umur 18th. Pinggang b...	Halo Yeni. Terima kasih sudah bertanya. Keluha...	batu-ginjal	[F1]. Halo Yeni. Terima kasih sudah bertanya. ...	NaN
248	498	Pusing, mual, sakit saat berkemih, dan keringa...	Sudah 2 hari ini saya mrasakn pusing,mual dise...	Halo. Gejala mual, berkeringat dingin dengan r...	batu-ginjal	[F1]. Halo. [F4]. Gejala mual, berkeringat din...	NaN
249	499	Perut sakit sekali setelah buang air kecil	Dok, Mau tanya nih. Saya itu kHalo habis bua...	Hai. Nyeri perut setelah buang air kecil dapat...	batu-ginjal	[F1]. Hai. [F3]. Nyeri perut setelah buang air...	NaN

500 rows x 7 columns

*Gambar 7 gabungan Dataset*

- **Membuang Tabel yang tidak diperlukan**

Pada kode sumber 4.4 akan dilakukan membuang(drop) Tabel yang tidak diperlukan. Sampai menyisakan Tabel yang berisikan “no”, “answer”, “topic\_set”:

```

1. dataset =
   data.drop(['title', 'question', 'answer_labeled', 'labeled'], axis=1)
2. dataset = dataset.reset_index()
3. dataset = dataset.drop(['index'], axis=1)
4. dataset

```

Kode sumber 4.4 Membuang Tabel yang tidak diperlukan

No_Id	answer	topic_set
0	0 Halo yayan, Terima kasih atas pertanyaannya. B...	Gangguan Pencernaan
1	1 Halo. Terima kasih atas pertanyaan yang disamp...	Intim Wanita
2	2 Selamat pagi, terima kasih atas pertanyaannya....	Obat
3	3 Halo Friscilla, terima kasih ya sudah bertanya...	Gangguan Pencernaan
4	4 Halo, terima kasih atas pertanyaannya untuk Al...	Asam Lambung
...	...	...
495	495 Halo Anatasya..... Terima kasih atas pertany...	batu-ginjal
496	496 Hai. Keluhan tersebut dapat disebabkan oleh: i...	batu-ginjal
497	497 Halo Yeni. Terima kasih sudah bertanya. Keluha...	batu-ginjal
498	498 Halo. Gejala mual, berkeringat dingin dengan r...	batu-ginjal
499	499 Hai. Nyeri perut setelah buang air kecil dapat...	batu-ginjal

500 rows x 3 columns

Gambar 8 Membuang Tabel yang tidak diperlukan

### 4.3.2 Segmentasi Teks

Data dari answer akan dilakukan *split* perkalimat dengan menggunakan metode Segmentasi Teks. Metode ini juga bisa

memilah atau membagi perkalamat menjadi satu kelompok secara otomatis dengan arahan dari komputer itu sendiri.

- **Lowercase dataset**

Di kode sumber 4.5 baris 1-4 dataset diubah menjadi huruf kecil semua(lowerces).

```
1. dataset['answer'] =  
   dataset['answer'].str.lower()  
2. dataset = dataset.reset_index()  
3. dataset = dataset.drop(['index'], axis=1)  
4. dataset = dataset.astype({'No_Id': 'int'})
```

*Kode sumber 4.5 lowercase*

- **Split dataset**

Selanjutnya kode sumber 4.6 baris 1-3 data akan di split perkalamat dengan pembatasnya (.).

```
1. for index, data in  
   enumerate(dataset['answer'][0:500]):  
2.   datasplit = data.split('.')  
3.   datasplit = list(filter(None, datasplit))
```

*Kode sumber 4.6 Split dataset*

- **Proses Segment teks**

Kemudian kode sumber 4.7 baris 8-15 Bersama dengan dataset di split akan di proses untuk memberi value segment di setiap data yang di split.

```
8.   sentenced_text = datasplit  
9.   vecr =  
   CountVectorizer(vocabulary=wrddvecs.index)
```

```

10. if(len(datasplit) <= 3):
11.     continue
12.     sentence_vectors =
        vecr.transform(sentenced_text).dot(wrdvecs)
13.     penalty = get_penalty([sentence_vectors],
        segment_len)
14.     optimal_segmentation =
        split_optimal(sentence_vectors, penalty,
        seg_limit=250)
15.     segmented_text =
        get_segments(sentenced_text,
        optimal_segmentation)

```

*Kode sumber 4.7 Proses Segment teks*

- **Membuat table Segment teks**

Kode sumber 4.8 baris 20-35 table ini akan menempatkan value hasil dari dataset yang displit(No\_ID) dan value segment(segment).

```

16. segment_array = []
17. for i in range(0, len(segmented_text)):
18.     for sentence in segmented_text[i]:
19.         segment_array.append(i)
20.     if(index ==0):
21.         df = pd.DataFrame()
22.         df['buangan'] = datasplit
23.         df['No_ID'] = dataset['No_Id'][index]
24.         df['Answer'] = datasplit
25.         df.pop("buangan")
26.         df['segment'] = segment_array
27.     else:

```

```

28.     temp = pd.DataFrame()
29.     temp['buangan'] = datasplit
30.     temp['No_ID'] = dataset['No_Id'][index]
31.     temp['Answer'] = datasplit
32.     temp.pop("buangan")
33.     temp['segment'] = segment_array
34.     df = df.append(temp)
35.         df = df.reset_index(drop=True)

```

*Kode sumber 4.8 Membuat table Segment teks*

- **Memberi topic\_set pada hasil Segment teks**

Kode sumber 4.9 baris 1-4 mengambil dataset dan membuang table answer. baris 5-7 untuk menggabungkan topic\_set dari dataset pada hasil dari segmentasi.

```

1. dataset =
   pd.read_excel('/content/drive/MyDrive/TA &
   KP/Dataset/dataset.xlsx')
2. ds = dataset.drop(['answer'], axis=1)
3. # ds = ds.drop(['index'], axis=1)
4. ds = ds.rename(columns={'No_Id': 'No_ID'})

5. segment = df.merge(ds, on="No_ID")
6. segment=segment[['No_ID', 'topic_set', 'Answer',
   'segment']]
7. segment

```

*Kode sumber 4.9 Memberi topic\_set*

No_ID	Answer	segment	topic_set
0	0	halo yayan, terima kasih atas pertanyaannya	0 Gangguan Pencernaan
1	0	berdasarkan riwayat keluhan istri anda yang s...	1 Gangguan Pencernaan
2	0	gejala muntah dan diare pada dasarnya merupak...	1 Gangguan Pencernaan
3	0	kondisi ini dapat membaik dengan sendirinya d...	2 Gangguan Pencernaan
4	0	pantau terus gejala yang muncul, kembalikan c...	2 Gangguan Pencernaan
...	...	...	...
8576	499	hindari memijat / mengurut bagian perut yang ...	3 batu-ginjal
8577	499	jika nyeri terasa mengganggu dapat mengonsums...	3 batu-ginjal
8578	499	baca laman artikel berikut ini: nyeri saat bu...	3 batu-ginjal
8579	499	demikian semoga bermanfaat	3 batu-ginjal
8580	499	salam, dr muliani sukiman	3 batu-ginjal

8581 rows x 4 columns

*Gambar 9 Hasil segmentasi*

- **Memberi Id\_kalimat**

Kode sumber 4.10 ini akan memberikan Id\_kalimat pada setiap paragraph yang di split.

```

1. a = 0
2. count = 1
3. segment['id_kalimat'] = None
4.
5. for i in range(len(segment)):
6.     # print(i)
7.     if(a == segment['No_ID'].iloc[i]):
8.         segment['id_kalimat'].iloc[i] = count
9.         count = count+1
10.
11.    else:
12.        count = 0
13.        segment['id_kalimat'].iloc[i] = count
14.        count = count+1

```

```

15.     a = a+1
16.
17. segment

```

*Kode sumber 4.10 Memberi Id\_kalimat*

No_ID	Topic_set	Answer	segment	id_kalimat	
0	0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya	0	1
1	0	ganggu cerna	berdasarkan riwayat keluhan istri anda yang s...	1	2
2	0	ganggu cerna	gejala muntah dan diare pada dasarnya merupak...	1	3
3	0	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	2	4
4	0	ganggu cerna	pantau terus gejala yang muncul, kembalikan c...	2	5
...	...	...	...	...	...
8576	499	batu ginjal	hindari memijat / mengurut bagian perut yang ...	3	9
8577	499	batu ginjal	jika nyeri terasa mengganggu dapat mengonsums...	3	10
8578	499	batu ginjal	baca laman artikel berikut ini: nyeri saat bu...	3	11
8579	499	batu ginjal	demikian semoga bermanfaat	3	12
8580	499	batu ginjal	salam, dr muliani sukiman	3	13

8581 rows × 5 columns

*Gambar 10 Memberi Id\_kalimat*

### 4.3.3 Praproses

Dari dataset akan dibuat daftar kata sebagai penentu setiap label dari hasil segmentasi. Membuat daftar kata diperlu untuk membersihkan kata-kata yang tidak perlu. Data dari answer juga akan di pilah-pilah menjadi perkata, sehingga kita bisa menganalisa tiap “label” dan “topic\_set” dari data Fungsi Wawancara. Selanjutnya perkata akan disusun dari frekuensi yang paling banyak muncul sebagai penentu dari label tersebut.

- **Memuat Data hasil wawancara**

Pada langkah ini, dapat memuat data dengan kode sumber 4.11 dari baris 1-2, yaitu file “Data Gabung 2-1.xlsx dan Data Gabung



2-2.xlsx”. Kemudian kedua data akan digabungkan dalam baris 3-6. Selanjutnya beberapa table akan di rename dalam baris 7. Di baris 8-10 akan diberikan table topic\_set.

```
1. dafla1 =
   pd.read_excel('/content/drive/MyDrive/TA &
   KP/Data wawancara/Data Gabung 2-1.xlsx')
2. dafla2 =
   pd.read_excel('/content/drive/MyDrive/TA &
   KP/Data wawancara/Data Gabung 2-2.xlsx')

3. dafla2['No Urut'] = dafla2['No Urut'] + 4021
4. dafla2['Id Data'] = dafla2['Id Data'] + 250
5. dafla = [dafla1, dafla2]
6. dafla = pd.concat(dafla)
7. dafla = dafla.rename(columns={'Id Data':
   'No_ID', 'Data Indonesia': 'answer'})

8. dafla=dafla[['No_ID', 'answer', 'Label']]
9. dafla = dafla.merge(ds, on="No_ID")
10. dafla=dafla[['No_ID', 'topic_set', 'answer', 'Label']]
11. dafla
```

*Kode sumber 4.11 Memuat Daftar Kata*

No_ID	Topic_set	answer	Label	id_kalimat
0	0 ganggu cerna	Halo yayan, Terima kasih atas pertanyaannya	F1	1
1	0 ganggu cerna	Berdasarkan riwayat keluhan istri Anda yang su...	F4	2
2	0 ganggu cerna	Gejala muntah dan diare pada dasarnya merupaka...	F3	3
3	0 ganggu cerna	Kondisi ini dapat membaik dengan sendirinya da...	F3	4
4	0 ganggu cerna	Pantau terus gejala yang muncul, kembalikan ca...	F5	5
...	...	...	...	...
8569	499 batu ginjal	Hindari memijat / mengurut bagian perut yang...	F5	9
8570	499 batu ginjal	Jika nyeri terasa mengganggu dapat mengonsumsi...	F5	10
8571	499 batu ginjal	Baca laman artikel berikut ini: Nyeri saat bua...	F5	11
8572	499 batu ginjal	Demikian semoga bermanfaat	F1	12
8573	499 batu ginjal	Salam, dr Muliiani Sukiman	F1	13

8574 rows × 5 columns

*Gambar 11 Hasil Membuat Daftar Kata*

- ***Preprocessing***

Kode sumber 4.12 baris 1 akan dibuat table baru “ans\_clean”. baris 2 akan menampilkan table dafla dalam variable baru. baris 3-5 fungsi yang akan dipakai dalam melakukan preprocessing. Baris 6 data answer akan di lowercase (semua huruf menjadi kecil). Baris 7-10 semua data ans\_clean akan dibersihkan dan akan dipilah-pilah berdasarkan perkata yang sudah menjadi kata dasar.

```

1. dafla['ans_clean'] = ''
2. preproc = dafla[['No_ID', 'Topic_set', 'answer',
  'Label', 'id_kalimat', 'ans_clean']]
3. lemmatizer = Lemmatizer()
4. stop_words = stopwords.words("indonesian")
5. word_detokenizer = TreebankWordDetokenizer()

```

```
6. prepoc['ans_clean'] =
    prepoc['answer'].str.lower()
7. prepoc['ans_clean'] =
    prepoc['ans_clean'].apply(lambda s:
    re.sub(r'^\w\s', '', re.sub(r'\d+', '', s)))
8. prepoc['ans_clean'] =
    prepoc['ans_clean'].apply(lambda s:
    lemmatizer.lemmatize(s))
9. prepoc['ans_clean'] =
    prepoc['ans_clean'].apply(lambda s: [v for v
    in s.split() if v not in stop_words])
10. prepoc['ans_clean'] =
    prepoc['ans_clean'].apply(lambda s:
    word_detokenizer.detokenize(s))
11. prepoc
```

*Kode number 4.12 Preprocessing*

	No_ID	Topic_set	answer	Label	id_kalimat	ans_clean
0	0	ganggu cerna	Halo yayan, Terima kasih atas pertanyaannya	F1	1	halo yayan terima kasih
1	0	ganggu cerna	Berdasarkan riwayat keluhan istri Anda yang su...	F4	2	dasar riwayat keluh istri muntah muntah diare ...
2	0	ganggu cerna	Gejala muntah dan diare pada dasarnya merupaka...	F3	3	gejala muntah diare dasar rupa respon tubuh wa...
3	0	ganggu cerna	Kondisi ini dapat membaik dengan sendirinya da...	F3	4	kondisi hitung obat khusus dokter
4	0	ganggu cerna	Pantau terus gejala yang muncul, kembalikan ca...	F5	5	pantau gejala muncul cair tubuh monitor tanda ...
...	...	...	...	...	...	...
8569	499	batu ginjal	Hindari memijat / mengurut bagian perut yang...	F5	9	hindar pijat urut perut sakit
8570	499	batu ginjal	Jika nyeri terasa mengganggu dapat mengonsumsi...	F5	10	nyeri ganggu konsumsi paracetamol ibuprofen reda
8571	499	batu ginjal	Baca laman artikel berikut ini: Nyeri saat bua...	F5	11	baca laman artikel nyeri buang air infeksi sal...
8572	499	batu ginjal	Demikian semoga bermanfaat	F1	12	moga manfaat
8573	499	batu ginjal	Salam, dr Muliani Sukiman	F1	13	salam dr muliani sukiman

8574 rows × 6 columns

*Gambar 12 Preprocessing*

- **Membuat Daftar Kata**

Kode sumber 4.13 baris 1 akan dibuat table F1 sampai F6. Kode sumber 2 menggunakan fungsi *unique* agar Topic\_set yang sama akan di jadikan 1. Baris 3-4 loop untuk membedakan tiap *topic* dan *label*. Baris 5 akan membuat data dengan nama (*topic*)\_(*label*).txt. Baris 6-7 untuk mengiterasi setiap baris pada data preproc yang memiliki nilai row[topic\_set] sama dengan loop topic dan row[Label] sama dengan loop label.

```
1. Labels = ['F1', 'F2', 'F3', 'F4', 'F5', 'F6']
2. topic_set = preproc['Topic_set'].unique()
3. for topic in topic_set:
4. for label in labels:
5. f =
    open(f"/content/drive/MyDrive/TA/dafka/{topic}
    _{label}.txt", 'w')
6. for _, row in preproc.iterrows():
7. if row['Topic_set'] == topic and row['Label']
    == label:
            f.write(row['ans_clean'] + "\n")
8. f.close()
```

Kode sumber 4.13 Membuat Daftar Kata

- **Membuat Daftar Frequency**

Kode sumber 4.14 baris 1-3 loop untuk membaca data berdasarkan (*topic*)\_(*label*). Baris 4-6 menghapus karakter whitespace dari setiap kalimat, tiap kata yang yang memiliki hanya 1 huruf tidak akan di baca, tiap kata yang sama akan di jadikan 1.

Baris 7-10 perulangan sentence agar bisa menjumlahkan tiap kata yang sama dari fungsi *unique*. Baris 11 membuat dataframe untuk menempatkan daftar frequency. Baris 12 mengurutkan dari yang terbanyak dari daftar frequency. Baris 13 membuat data daftar frequency.

```
1. for topic in topic_set:
2. for label in labels:
3. with
   open(f"/content/drive/MyDrive/TA/dafka/{topic}
      _{label}.txt") as f:
       sentences = f.readlines()
4. sentences = [sentence.strip() for sentence in
   sentences]
5. unique_words = set([word for sentence in
   sentences for word in sentence.split() if
   len(word) > 1])
6. doc_freq = {word: 0 for word in unique_words}

7. for sentence in sentences:
   words_in_sentence = set(sentence.split())
8. for word in unique_words:
9. if word in words_in_sentence:
   doc_freq[word] += 1
10.doc_freq = {
   'word': doc_freq.keys(),
   'doc_freq': doc_freq.values(),
   }
11.doc_freq = pd.DataFrame(doc_freq)
12.doc_freq = doc_freq.sort_values('doc_freq',
   ascending=False)
```

```
13.doc_freq.to_csv(f"/content/drive/MyDrive/TA/da
fka//DF_{topic}_{label}.csv", index=False)
```

*Kode sumber 4.14 Membuat Frequency*

### 4.3.4 Wordcloud

Dari datafrequency akan divisualisasi hasil dari daftar frequency ang sering muncul di tiap topic dan label.

- **Filter topic dan label untuk**

Kode sumber 4.15 baris 1 filter data prepoc berdasarkan Topic\_set dan Label. baris 2-4 akan membersihkan kata dari table “ans\_clean”. baris 6-7 untuk menghitung frequency dari tiap kata yang sama.

```
1. ts = prepoc[(prepoc.Topic_set == "asam
lambung") & (prepoc.Label == "F1")]
2. a = ts['ans_clean'].str.lower().str.cat(sep='
')
3. a = a.replace("\n", " ") # remove line break
4. a = re.sub(r'^[a-zA-Z]', ' ', a) #remove
symbol dan angka
5. words = nltk.tokenize.word_tokenize(a)
6. word_dist = nltk.FreqDist(words)
7. rslt = pd.DataFrame(word_dist.most_common(),
                        columns=['Word', 'Frequency'])
```

*Kode sumber 4.15 Fungsi filter daftar frequency*



	Word	Frequency
0	moga	15
1	terima	12
2	kasih	12
3	dr	12
4	halo	9
5	manfaat	7
6	salam	7
7	alodokter	6
8	informasi	6
9	bantu	6
10	dear	4

Gambar 13 hasil fungsi filter daftar frequency

- **Menampilkan hasil daftar frequency**

Kode sumber 4.16 baris 1-3 menggabungkan data berdasarkan table words dan menggunakan fungsi library wordcloud. Baris 4 untuk membuat ukuran gambar yang akan di tampilkan. Baris 5 untuk menampilkan wordcloud dalam bentuk gambar. Baris 6 untuk menghilangkan garis sumbu x dan y. Baris 7 untuk menampilkan hasil word cloud.

```

1. dataspt = ' '.join(words)
2. stopword = set(STOPWORDS)
3. wordcloud = WordCloud(stopwords=stopword,
    background_color='white').generate(dataspt)

```





menambahkan topic\_set berdasarkan No\_ID. Baris 8 membuat file baru.

```
1. segment_label = {
    'doc_id': [],
    'segment_id': [],
    'segment_text': [],
}
2. for doc_id in range(500):
    seg_ids = segment.loc[segment['No_ID'] ==
doc_id, 'segment'].unique()
3. for seg_id in seg_ids:
    seg_text = segment.loc[(segment['No_ID'] ==
doc_id) & (segment['segment'] == seg_id),
    'Answer'].to_list()
    seg_text = ' '.join(seg_text)
    segment_label['doc_id'].append(doc_id)
    segment_label['segment_id'].append(seg_id)
    segment_label['segment_text'].append(seg_text)
4. segment_label = pd.DataFrame(segment_label)

5. segment_label =
    segment_label.rename(columns={'doc_id':
    'No_ID', 'segment_id': 'segment',
    'segment_text': 'Answer'})
6. segment_label = segment_label.merge(ds,
on="No_ID")
7. segment_label=segment_label[['No_ID', 'Topi
c_set', 'Answer', 'segment']]
```

```
8. segment_label.to_excel('/content/drive/MyDrive
/TA/segmentasi/segment_label.xlsx',
index=False)
```

*Kode sumber 4.17 Menyatukan data berdasarkan segment*

No_ID	Topic_set	Answer	segment
0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya b...	0
1	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	1
2	ganggu cerna	selain itu, berikut anjuran yang dapat dilaku...	2
3	ganggu cerna	minum air, sup, atau minuman elektrolit untuk...	3
4	ganggu cerna	menghindari susu dan produk susu, kopi, minum...	4
...	...	...	...
2982	batu ginjal	hai nyeri perut setelah buang air kecil dapat...	0
2983	batu ginjal	dengan demikian penanganan yang tepat dapat d...	1
2984	batu ginjal	perbanyak konsumsi air putih 2-3 liter sehari...	2
2985	batu ginjal	hindari memijat / mengurut bagian perut yang ...	3
2986	batu ginjal	jika nyeri terasa mengganggu dapat mengonsums...	4

2987 rows × 4 columns

*Gambar 15 Menyatukan data berdasarkan segment*

- **Memprediksi label berdasarkan segment**

Kode sumber 4.18 baris 1-3 membuat distance kosong berdasarkan topic dan label yang akan dibaca dalam file yang sudah dibuat. Baris 4 membuat label(F1-F6) menjadi nilai awal 0. Baris 5 untuk mengiterasi setiap baris pada data seg\_clean yang memiliki nilai row[ans\_clean\_set]. Baris 6-7 pengulangan data berdasarkan kata-kata yang ada dalam baris dan dalam 1 topic.

Baris 8-9 dalam 1 topic yang sama diceknya tiap kata, jika ada kata yang sama ada maka tiap label bagian kata yang sama akan bertambah 1. Baris 10-13 tiap label akan dicari nilai yang tertinggi kemudian dalam table predicted\_label akan dicari nilai tertinggi sehingga bisa menentukan labelnya.

```
1. df_dict = {}
2. for topic in topic_set:
    df_dict[topic] = {}
3. for label in labels:
    df_dict[topic][label] =
    pd.read_csv(f"/content/drive/MyDrive/TA/dafka/
    DF_{topic}_{label}.csv")
4. for label in labels:
    seg_clean[label] = 0
5. for i, row in seg_clean.iterrows():
    sentence = row['ans_clean'].split()
6. for word in sentence:
7. for topic in topic_set:
8. if row['Topic_set'] != topic: continue
9. for label in labels:
    if
    df_dict[topic][label]['word'].isin([word]).any
    ():
    seg_clean.loc[i, label] += 1
10. f_cols = seg_clean.loc[:, 'F1':'F6'].columns
11. max_cols = seg_clean[f_cols].idxmax(axis=1)
12. max_counts = seg_clean[f_cols].max(axis=1)
```

```
13. seg_clean['predicted_label'] =  
    np.where(max_counts.eq(max_counts.groupby(level=0).transform('max')), max_cols, 'Fx')
```

*Kode sumber 4.18 Prediksi label berdasarkan segment*

No_ID	Topic_set	Answer	ans_clean	segment	F1	F2	F3	F4	F5	F6	predicted_label	
0	0	ganggu cerna	halo yayan, terima kasih atas pertanyaannya b...	halo yayan terima kasih dasar riwayat keluh is...	0	4	9	33	32	30	1	F3
1	0	ganggu cerna	kondisi ini dapat membaik dengan sendirinya d...	kondisi hitung obat khusus dokter pantau gejala...	1	0	9	21	16	28	1	F5
2	0	ganggu cerna	selain itu, berikut anjuran yang dapat dilaku...	anjur laku keluh hindar makan jam	2	0	1	5	4	6	1	F5
3	0	ganggu cerna	minum air, sup, atau minuman elektrolit untuk...	minum air sup minum elektrolit bantu cair tubu...	3	1	3	12	9	19	0	F5
4	0	ganggu cerna	menghindari susu dan produk susu, kopi, minum...	hindar susu produk susu kopi minum soda makan ...	4	8	3	21	15	25	1	F5
...	...	...	...	...	...	...	...	...	...	...	...	...
2982	499	batu ginjal	hai nyeri perut setelah buang air kecil dapat...	hai nyeri perut buang air infeksi salur kemih ...	0	7	16	33	28	32	12	F3
2983	499	batu ginjal	dengan demikian penanganan yang tepat dapat d...	tangan sesuai kondisi dasar anjur istirahat ko...	1	0	2	13	9	13	0	F3
2984	499	batu ginjal	perbanyak konsumsi air putih 2-3 liter sehari...	konsumsi air putih liter tahan buang air buang...	2	0	7	10	9	10	5	F3
2985	499	batu ginjal	hindari memijat / mengurut bagian perut yang ...	hindar pijat urut perut sakit	3	2	1	5	4	5	0	F3
2986	499	batu ginjal	jika nyeri terasa mengganggu dapat mengonsums...	nyeri ganggu konsumsi paracetamol ibuprofen re...	4	9	7	16	11	18	8	F5

2987 rows x 12 columns

*Gambar 16 Prediksi label berdasarkan segment*

### 4.3.6 Cosine Similarity

Pada tahap ini akan dicari nilai kemiripan antara hasil prediksi dari segment dengan data yang didapat dari hasil wawancara medis.

- **Perhitungan Cosine Similarity**

Kode sumber 4.19 baris 1 membuat table baru untuk cosine similarity pada data seg\_clean. baris 3 membaca data label yang sudah dibuat dengan syarat 1 huruf kata tidak bisa ikut kepanjang. baris 4 untuk mengiterasi setiap baris pada data seg\_clean dengan pengaturan row[predicted\_label] sama dengan label. Baris 5-7 melakukan perhitungan *Cosine Similarity*. Baris 8 untuk mengiterasi setiap baris pada data seg\_clean dengan pengaturan row[predicted\_label] sama dengan label dan memasukan nilai hasil *Cosine Similarity*. Baris 9-14 akan memanggil fungsi addCosineSimilarity, yang memiliki nilai F1 dari label. Baris 15 untuk memanggil nilai cosine\_similarity yang lebih dari nilai x. Baris 16 untuk menampilkan grafik dari perhitungan cosine\_similarity.

```
1. seg_clean['cosine_similarity'] = ''
2. def addCosineSimilarity(label1: str):
3. f =
   open(f"/content/drive/MyDrive/TA/dafka/{label1}
      }.txt")
   sentences = [s.strip() for s in f.readlines()]
   sentences = ["
   ".join(list(itertools.chain.from_iterable(list
      (filter(lambda s: len(s) > 1, [s.split() for s
      in sentences])))))]
```

```

4. for _, row in seg_clean.iterrows():
    if row['predicted_label'] == label1:
        sentences.append(row['ans_clean'])
5. vectorizer = CountVectorizer()
6. x = vectorizer.fit_transform(sentences)
7. similarities =
    pd.DataFrame(cosine_similarity(x))
    i = 1
8. for index, row in seg_clean.iterrows():
    if row['predicted_label'] == label1:
        seg_clean.loc[index, 'cosine_similarity'] =
            similarities[0][i]
            i += 1
9. addCosineSimilarity('F1')
10. addCosineSimilarity('F2')
11. addCosineSimilarity('F3')
12. addCosineSimilarity('F4')
13. addCosineSimilarity('F5')
14. addCosineSimilarity('F6')

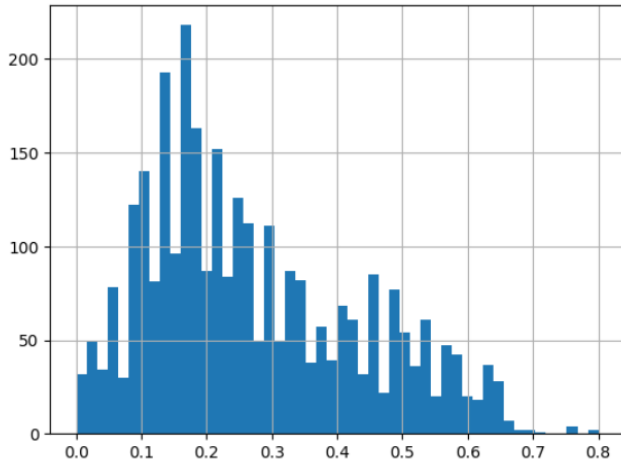
15. seg_clean['is_label_correct'] =
    seg_clean['cosine_similarity'].apply(lambda x:
    1 if x >= 0.45 else 0)
16. seg_clean.cosine_similarity.astype(float).round(
    2).hist(bins=40)

```

*Kode sumber 4.19 Perhitungan cosine*



<Axes : >



*Gambar 17 grafic perhitungan cosine*

*[Halaman ini sengaja dikosongkan]*

## BAB V PENGUJIAN DAN EVALUASI

Bab ini menjelaskan tahap uji coba terhadap pelabelan segmentasi teks dengan data tanya-jawab alodoc. Pengujian dilakukan untuk memastikan fungsionalitas dan kesesuaian hasil implementasi segmentasi teks.

### 5.1. Tujuan Pengujian

Pengujian dilakukan terhadap data tanya-jawab alodoc guna melabelkan tanya-jawab dari alodoc berdasarkan fungsi wawancara medis.

### 5.2. Skenario Pengujian

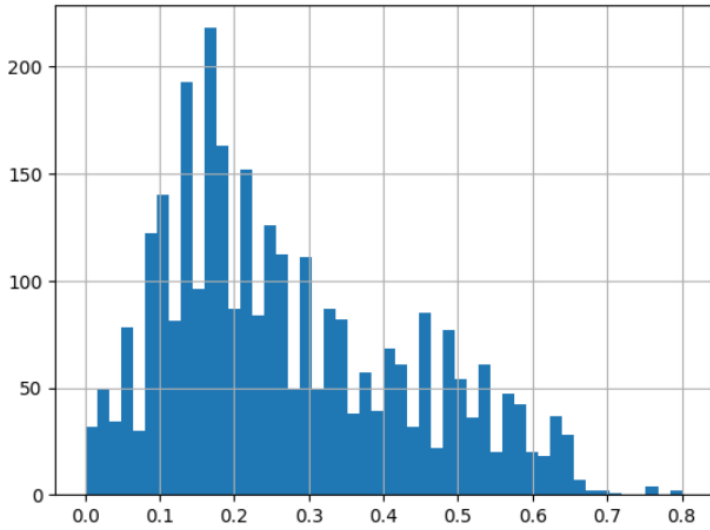
Skenario mencoba perubahan hasil segmentasi teks ke pelabelan berdasarkan daftar frequency kata yang di dapat dari faktor fungsi wawancara medis.

### 5.3. Hasil dan Evaluasi Pengujian

Hasil evaluasi pelabelan dari segmentasi teks akan berbeda dari fungsi faktor wawancara medis dengan menggunakan *Cosine Simeilarity*. Berdasarkan *threshold* 0.25 keatas dengan nilai 1 masuk dalam kemiripan. *Threshold* yang rendah dikarenakan dalam pelabelan hasil segmentasi berdasarkan daftar kata yang sering keluar pada pengelompokan tiap segmentnya.

```
print(counts)
0    1655
1    1381
Name: is_label_correct, dtype: int64
```

Gambar 18 hasil threshold 0.25 keatas



*Gambar 19 hasil total Cosine Similarity*

## **BAB VI**

### **KESIMPULAN DAN SARAN**

Bab ini menjelaskan tahap uji coba terhadap pelabelan berdasarkan hasil segmentasi teks.

#### **6.1. Kesimpulan**

Sebelum dilakukan Segmentasi Teks, data tanya jawab alodokter akan dipilah menjadi beberapa kalimat.

Dalam melakukan Segmentasi Teks data tanya jawab alodokter akan melakukan proses representation, scoring dan splitting sehingga mendapatkan nilai pada tiap segment yang berkohorensi tinggi. Selanjutnya hasil dari Segmentasi Teks akan dipra-proses. Hasil pra-proses Segementasi Teks akan dibuat data unique segment untuk dibandingkan dengan hasil data Daftar Frekuensi sehingga mendapatkan prediksi label pada setiap segment.

Setelah melakukan pelabelan berdasarkan hasil segmentasi teks pada kegiatan kerja praktek adalah masih kurangnya akurasi yang sesuai berdasarkan faktor fungsi wawancara medis.

#### **6.2. Saran**

Saran untuk perancangan pelabelan berdasarkan hasil segmentasi adalah perbanyak daftar frequency tiap label dan jangan memberi Batasan pada daftar frequency.

*[Halaman ini sengaja dikosongkan]*

## DAFTAR PUSTAKA

- Alemi, A. A., & Ginsparg, P. (2015, Mar 18). Text Segmentation based on Semantic Word Embeddings.  
doi:<https://doi.org/10.48550/arXiv.1503.05543>
- Andana, N. D. (2021). *PENCARIAN JUDUL BUKU MENGGUNAKAN PENDEKATAN COSINE SIMILARITY DAN EUCLIDEAN DISTANCE*. Yogyakarta: Universitas Sanata Dharma Yogyakarta.  
Retrieved from  
[https://repository.usd.ac.id/40483/2/155314063\\_full.pdf](https://repository.usd.ac.id/40483/2/155314063_full.pdf)
- Chen, S., Guo, X., Wu, T., & Ju, X. (2020). Information Processing and Management.  
doi:<https://doi.org/10.1016/j.ipm.2020.102253>
- DPTSI. (n.d.). Retrieved from ITS:  
<https://www.its.ac.id/informatika/id/tentang-kami/visi-dan-misi/>
- Dr. Ir. Siti Rochimah MT., P. (n.d.). *LABORATORIUM REKAYASA PERANGKAT LUNAK*. Retrieved from ITS:  
<https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-rekayasa-perangkat-lunak/>
- Dr. Wahyu Suadi, S. M. (n.d.). *LABORATORIUM ARSITEKTUR DAN JARINGAN KOMPUTER*. Retrieved from ITS:  
<https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-arsitektur-dan-jaringan-komputer/>
- Dr. Yudhi Purwananto, S. M. (n.d.). *Dr. Yudhi Purwananto, S.Kom., M.Kom.* Retrieved from ITS:  
<https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-algoritma-dan-pemrograman/>

- Dr.Eng. Chastine Faticah, S. M. (2020, 1 2). Retrieved from ITS: <https://www.its.ac.id/informatika/id/tentang-kami/sambutan-kepala-departemen/>
- Efrizoni, L., Defit, S., Tajuddin, M., & Anggrawan, A. (2020). *Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning*. 21. doi:<https://doi.org/10.30812/matrik.v21i3.1851>
- Hagiwara, M. (2021). *Real-World Natural Language Processing*. New York: Manning Publications Co.
- Imam Kuswardayan, S. M. (n.d.). *LABORATORIUM GRAFIKA INTERAKSI DAN GAME*. Retrieved from ITS: <https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-grafika-interaksi-dan-game/>
- Ma, A. K., & Hoppe, R. B. (2013, September). “Best Practice” for Patient-Centered Communication: A Narrative Review. *Journal of Graduate Medical Education*. doi:<https://doi.org/10.4300/JGME-D-13-00072.1>
- Oghbaie, M., & Zanjireh, M. M. (2018). *Pairwise document similarity measure based on present term set*. doi:<https://doi.org/10.1186/s40537-018-0163-2>
- Prof. Dr.Eng. Nanik Suciati, S. (n.d.). *LABORATORIUM KOMPUTASI CERDAS DAN VISI*. Retrieved from ITS: <https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-komputasi-cerdas-dan-visi/>
- Prof. Drs. Ec. Ir. Riyanarto Sarno, M. P. (n.d.). *LABORATORIUM MANAJEMEN CERDAS INFORMASI*. Retrieved from ITS: <https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-manajemen-cerdas-informasi/>
- Prof. Tohari Ahmad, S. P. (n.d.). *LABORATORIUM KOMPUTASI BERBASIS JARINGAN*. Retrieved from



ITS:

<https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-komputasi-berbasis-jaringan/>

- Purnat, T. (2020). Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 22, 1. doi:<https://doi.org/10.2196/19659>
- Rahutomo, F., Ikawati, D. S., & Rohman, O. A. (2019, Juni). EVALUASI FITUR WORD2VEC PADA SISTEM UJIAN ESAI ONLINE. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 4, 36-45. doi:<https://doi.org/10.29100/jipi.v4i1.1098.g588>
- Saputra, W. S., & Muttaqin, F. (2013, September 18). PENGENALAN KARAKTER PADA PROSES DIGITALISASI DOKUMEN MENGGUNAKAN COSINE SIMILARITY. *SANTIKA*, 51-56. Retrieved from <http://eprints.upnjatim.ac.id/id/eprint/4770>
- Victor Hariadi, S. M. (n.d.). *LABORATORIUM PEMODELAN DAN KOMPUTASI TERAPAN*. Retrieved from ITS: <https://www.its.ac.id/informatika/id/fasilitas/laboratorium/laboratorium-pemodelan-dan-komputasi-terapan/>
- Wang, J.-H., & Liu, T.-W. (2018). *An LSTM Approach to Short Text Sentiment Classification with*. Beijing, Chinese: University of Science and Technology Beijing.
- Yanuarti, R. (2021, 8 31). Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19. *JUSTINDO*, 6. doi:<https://doi.org/10.32528/justindo.v6i2.5503>

*[Halaman ini sengaja dikosongkan]*

## **BIODATA PENULIS I**

Nama : I Kadek Ricky Suirta  
Tempat, Tanggal Lahir : Kraksaan, 03 Oktober 1999  
Jenis Kelamin : Pria  
Telepon : +6281336276613  
Email : kadekricky03@gmail.com

### **AKADEMIS**

Kuliah : Departemen Teknik Informatika –  
FTEIC , ITS  
Angkatan : 2018  
Semester : 10 (sepuluh)