

**PENGEMBANGAN APLIKASI FITUR PREPROCESSING
UNTUK PEMROSESAN BAHASA ALAMI**

Disusun Sebagai Salah Satu Persyaratan dalam Memperoleh Gelar Sarjana (S-1)
pada Program Studi Teknik Informatika



Diajukan oleh:

MUHAMMAD TSAQIF ASSHIDDIQ

17106050031

**STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA
PROGRAM STUDI INFORMATIKA**

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2023

PENGESAHAN TUGAS AKHIR



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-54/Un.02/DST/PP.00.9/01/2024

Tugas Akhir dengan judul : Pengembangan Aplikasi Fitur Preprocessing Untuk Pemrosesan Bahasa Alami

yang dipersiapkan dan disusun oleh:

Nama : MUHAMMAD TSAQIF ASSHIDDIQ
Nomor Induk Mahasiswa : 17106050031
Telah diujikan pada : Selasa, 19 Desember 2023
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 659f5f6fcbba



Penguji I
Agus Mulyanto, S.Si., M.Kom., ASEAN Eng.
SIGNED

Valid ID: 659f56dded86



Penguji II
Kartiansmara Lilih Purnaumbara, M.Sc
SIGNED

Valid ID: 659be5da26149



Yogyakarta, 19 Desember 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 659f68d5eb463

SURAT PERNYATAAN KEASLIAN SKRIPSI

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Muhammad Tsaqif Asshiddiq

NIM : 17106050031

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa:

1. Skripsi yang berjudul: **Pengembangan Aplikasi Fitur Preprocessing Untuk Pemrosesan Bahasa Alami** merupakan hasil karya asli saya yang diajukan untuk memenuhi salah satu persyaratan memperoleh gelar sarjana strata satu (S1) di jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga, Yogyakarta.
2. Semua sumber yang saya gunakan dalam penulisan skripsi ini telah saya cantumkan sesuai dengan ketentuan yang berlaku, Jika di kemudian hari terbukti bahwa karya ini bukan hasil karya saya atau hasil plagiat dari orang lain, maka saya bersedia menerima sanksi yang berlaku.

Yogyakarta, 11 Desember 2023

Yang menyatakan,

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



Muhammad Tsaqif Asshiddiq

NIM: 17106050031

SURAT PERSETUJUAN SKRIPSI / TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Muhammad Tsaqif Asshiddiq
NIM : 17106050031
Judul Skripsi : Pengembangan Aplikasi Fitur Preprocessing Untuk Pemrosesan Bahasa Alami

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini saya berharap agar skripsi/tugas akhir tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya saya ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 11 Desember 2023
Pembimbing,

Dr. Agung Fatwanto, S.Si., M.Kom.

NIP. 19780106 200212 2 001

PENGEMBANGAN APLIKASI FITUR PREPROCESSING UNTUK PEMROSESAN BAHASA ALAMI

Muhammad Tsaqif Asshiddiq

17106050031

ABSTRAK

Dengan banyaknya pengguna internet saat ini, banyak data yang bisa didapatkan dengan mudah di internet. Kumpulan data ini dapat diproses dan dimanfaatkan untuk menghasilkan informasi baru. Salah satu caranya adalah dengan menggunakan pemrosesan bahasa alami atau *NLP*. Teks merupakan jenis data yang paling banyak dan umum digunakan. Namun data teks yang dikumpulkan biasanya masih berupa teks yang formatnya tidak terstruktur. Agar dapat diproses menggunakan pemrosesan bahasa alami, maka teks harus melalui tahap penyesuaian terlebih dahulu yang disebut dengan *text preprocessing*.

Dalam penelitian ini penulis bertujuan untuk membuat aplikasi untuk mempermudah tahapan *text preprocessing*. Aplikasi ini berbasis web yang dibuat menggunakan bahasa pemrograman Python dan *framework* Flask. Metode pengembangan aplikasi menggunakan *extreme programming*. Terdapat 15 fitur *preprocessing* yang dapat diimplementasikan yaitu *remove username, remove retweet, remove hashtag, remove url / http, remove punctuation, remove symbol, remove number, remove duplicate, remove slang, remove abbreviation, remove elongated character, lowercase, remove stopword, join case dan tokenizing*.

Dalam penelitian ini, aplikasi diuji dengan menggunakan pengujian *alpha* dan *beta*. Pengujian *alpha* dilakukan sendiri oleh penulis dengan hasil seluruh fitur *text preprocessing* berhasil diimplementasikan. Pengujian *beta* dilakukan dengan melibatkan 30 orang responden dengan 2 jenis pengujian yaitu pengujian fungsional dan pengujian non fungsional. Hasil dari pengujian fungsional menunjukkan bahwa 98,8% responden menjawab ya dan terdapat 1,2% responden yang menjawab tidak. Hasil dari pengujian non-fungsional menunjukkan bahwa 60,5% responden menjawab sangat setuju, 23,3% responden menjawab setuju, 14,4% responden menjawab netral, 1,6% responden menjawab tidak setuju dan 0% responden yang menjawab sangat tidak setuju. Dari hasil pengujian ini dapat diambil kesimpulan bahwa aplikasi berjalan dengan cukup baik dan dapat mempermudah *preprocessing*.

Kata Kunci : Pemrosesan Bahasa Alami, *NLP*, *Text preprocessing*

MOTTO

“Hidup adalah belajar tanpa akhir. Saat kau berhenti belajar, di saat itulah kau mati.”

“Uang bukan segalanya tapi segalanya butuh uang.”

“Alasan manusia terus berjuang untuk hidup adalah demi untuk mendapatkan sesuatu yang disebut kebahagiaan.”

“Lebih baik gagal setelah berusaha dari pada gagal tanpa adanya usaha.”

“Barang siapa yang bersungguh-sungguh dalam melakukan sesuatu pasti akan berhasil.”

“Jenius adalah 1% inspirasi dan 99% kerja keras.”

(Thomas Alva Edison)

“Semua orang jenius. Tetapi jika anda menilai seekor ikan dari kemampuannya memanjat pohon, ia akan menjalani seluruh hidupnya dengan percaya bahwa ia bodoh.”

(Albert Einstein)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamin. Segala puji bagi Allah. Saya ingin mengucapkan terima kasih banyak kepada seluruh pihak yang telah membantu dan mendukung saya dalam menyelesaikan penelitian ini. Untuk itu skripsi ini saya persembahkan kepada :

1. Kedua orang tua tercinta yang selalu senantiasa mendukung, membantu serta mendo'akan yang terbaik untuk anaknya.
2. Kedua adik tersayang yang selalu memberi semangat serta motivasi untuk terus berusaha.
3. Segenap keluarga serta saudara yang selalu memberikan saran serta nasehat.
4. Dr. Agung Fatwanto, S.Si., M.Kom. selaku Dosen Pembimbing Skripsi yang selalu sabar membimbing serta mengarahkan demi terselesaikannya penelitian ini.
5. Seluruh Dosen, staf dan karyawan di Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta, atas segala bantuan dan kerjasamanya selama ini.
6. Teman-teman seperjuangan Teknik Informatika 2017 yang sudah menemani kisah kuliah penuh perjuangan di kampus UIN Sunan Kalijaga.
7. Teman-teman komunitas, sahabat-sahabat serta kawan-kawan yang selalu membantu meskipun terhalang oleh jarak dan waktu.
8. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penelitian ini.

KATA PENGANTAR



Assalamu'alaikum Wr. Wb

Alhamdulillahirabbil'alamin. Segala puji bagi Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis bisa menyelesaikan penelitian yang berjudul **Pengembangan Aplikasi Fitur Preprocessing Untuk Pemrosesan Bahasa Alami**. Serta tidak lupa shalawat serta salam senantiasa dipanjatkan kepada Baginda Rasulullah Muhammad SAW yang selalu kita harapkan syafaatnya di Yaumul Qiyamah.

Dalam menyelesaikan penelitian ini, banyak halangan dan rintangan yang harus penulis hadapi. Namun pada akhirnya penelitian ini dapat terselesaikan dengan bantuan serta dukungan dari berbagai pihak yang senantiasa membantu baik secara langsung maupun tidak langsung. Untuk itu penulis ingin menyampaikan terima kasih kepada :

1. Kedua orang tua yang selalu mendukung, membimbing, serta memberi bantuan baik secara moral, spiritual, material dan selalu menjadi salah satu motivasi penulis untuk terus tetap berjuang.
2. Kedua adik yang selalu membantu dan memotivasi penulis untuk terus berusaha dan pantang menyerah.
3. Prof. Dr.Phil Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.

4. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
5. Maria Ulfa Siregar, S.Kom., MIT., Ph.D., selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta dan Dosen Pembimbing Akademik.
6. Dr. Agung Fatwanto, S.Si., M.Kom. selaku Dosen Pembimbing Skripsi yang selalu sabar membimbing serta mengarahkan demi terselesaikannya penelitian ini.
7. Seluruh Dosen, staf dan karyawan di Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta, atas segala bantuan dan kerjasamanya selama ini.
8. Seluruh keluarga besar yang selalu memberi dukungan, motivasi, saran, do'a dan bantuan yang tak terhitung jumlahnya.
9. Teman-teman seperjuangan Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
10. Teman-teman komunitas, sahabat-sahabat serta kawan-kawan yang selalu membantu meskipun terhalang oleh jarak dan waktu.
11. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu dalam penelitian ini.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu penulis mohon kepada pembaca untuk memberikan kritik, saran dan masukan yang membangun guna melengkapi dan memperbaiki kekurangan

dalam penyusunan penelitian ini. Semoga penelitian ini dapat bermanfaat bagi pembaca dan khususnya bagi penulis.

Yogyakarta, 13 Desember 2023

Penulis



Muhammad Tsaqif Asshiddiq

17106050031



DAFTAR ISI

PENGESAHAN TUGAS AKHIR	ii
SURAT PERNYATAAN KEASLIAN SKRIPSI.....	iii
SURAT PERSETUJUAN SKRIPSI / TUGAS AKHIR	iv
ABSTRAK	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR SOURCE CODE	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Batasan Masalah.....	5
1.5. Manfaat Penelitian.....	6
1.6. Keaslian Penelitian	6
1.7. Sistematika Penelitian	6
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	8
2.1. Tinjauan Pustaka	8
2.2. Landasan Teori	12
2.2.1. <i>Natural Language Processing</i>	12

2.2.2.	<i>Text preprocessing</i>	12
2.2.3.	Python	16
2.2.4.	Flask	17
2.2.5.	<i>Unified Modeling Language (UML)</i>	17
2.2.6.	<i>HTML</i>	20
BAB III METODE PENGEMBANGAN SISTEM		22
3.1.	Studi Pustaka	22
3.2.	Metodologi Pengembangan Sistem	22
3.2.1.	<i>Planning</i>	23
3.2.2.	<i>Design</i>	23
3.2.3.	<i>Coding</i>	24
3.2.4.	<i>Testing</i>	24
BAB IV ANALISIS DAN PERANCANGAN SISTEM		25
4.1.	Analisis Sistem	25
4.1.1.	Perangkat Pengembangan Sistem	25
4.1.2.	Analisis Kebutuhan Sistem	26
4.2.	Perancangan Sistem	28
4.2.1.	<i>Use Case Diagram</i>	29
4.2.2.	<i>Activity Diagram</i>	29
4.2.3.	<i>Interface</i>	31
BAB V IMPLEMENTASI DAN PENGUJIAN		33
5.1.	Implementasi Sistem	33
5.1.1.	Pengembangan <i>Front-End</i>	34
5.1.2.	Pengembangan <i>Back-End</i>	36
5.2.	Pengujian Sistem	48

5.2.1. Pengujian <i>Alpha</i>	48
5.2.2. Pengujian <i>Beta</i>	50
BAB VI HASIL DAN PEMBAHASAN	54
6.1. Hasil Pengembangan Sistem	54
6.1.1. Iterasi 1	54
6.1.2. Iterasi 2	54
6.1.3. Iterasi 3	55
6.2. Hasil Pengujian Sistem	56
6.2.1. Hasil Pengujian <i>Alpha</i>	56
6.2.2. Hasil Pengujian <i>Beta</i>	58
BAB VII PENUTUP	62
7.1. Kesimpulan	62
7.2. Saran	62
DAFTAR PUSTAKA	64

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	11
Tabel 2.2 Simbol Dalam <i>Use Case Diagram</i>	18
Tabel 2.3 Simbol Dalam <i>Activity Diagram</i>	19
Tabel 4.1 Informasi Perangkat Keras Pengembangan Sistem.....	25
Tabel 4.2 Informasi Perangkat Lunak Pengembangan Sistem.....	26
Tabel 5.1 Rencana Pengujian <i>Alpha</i>	49
Tabel 5.2 Skala <i>Likert</i>	51
Tabel 5.3 Rencana Pengujian Fungsional	52
Tabel 5.4 Rencana Pengujian Non-Fungsional	53
Tabel 6.1 Hasil Pengujian <i>Alpha</i>	56
Tabel 6.2 Hasil Pengujian Fungsional.....	58
Tabel 6.3 Hasil Pengujian Non-Fungsional	60

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 3.1 Diagram Model <i>Extreme Programming</i>	23
Gambar 4.1 <i>Use Case Diagram</i>	29
Gambar 4.2 <i>Activity Diagram</i>	30
Gambar 4.3 Desain <i>Interface</i> Halaman <i>Home</i>	31
Gambar 4.4 Desain <i>Interface</i> Halaman <i>Preprocessing</i>	32
Gambar 5.1 Struktur Direktori Aplikasi <i>Preprocessing</i>	33
Gambar 5.2 Desain Halaman <i>Home</i>	35
Gambar 5.3 Desain Halaman <i>Preprocessing</i>	35

DAFTAR SOURCE CODE

Source Code 5.1 Implementasi Upload File	37
Source Code 5.2 Implementasi Reset Fitur <i>Preprocessing</i>	38
Source Code 5.3 Implementasi <i>Download File</i>	38
Source Code 5.4 Implementasi Fitur <i>Remove Username</i>	39
Source Code 5.5 Implementasi Fitur <i>Remove Retweet</i>	39
Source Code 5.6 Implementasi Fitur <i>Remove Hashtag</i>	40
Source Code 5.7 Implementasi Fitur <i>Remove HTTP / URL</i>	40
Source Code 5.8 Implementasi Fitur <i>Remove Punctuation</i>	42
Source Code 5.9 Implementasi Fitur <i>Remove Symbol</i>	42
Source Code 5.10 Implementasi Fitur <i>Remove Number</i>	42
Source Code 5.11 Implementasi Fitur <i>Remove Duplicate</i>	43
Source Code 5.12 Implementasi Fitur <i>Remove Slang</i>	44
Source Code 5.13 Implementasi Fitur <i>Replace Abbreviation</i>	45
Source Code 5.14 Implementasi Fitur <i>Replace Elongated Character</i>	45
Source Code 5.15 Implementasi Fitur <i>Lower Case</i>	46
Source Code 5.16 Implementasi Fitur <i>Remove Stopword</i>	47
Source Code 5.17 Implementasi Fitur <i>Join Case</i>	47
Source Code 5.18 Implementasi Fitur <i>Tokenizing</i>	48

SUNAN KALIJAGA
YOGYAKARTA

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini jumlah pengguna internet di negara Indonesia sudah mencapai angka 215,62 juta jiwa dari total populasi 275,77 juta jiwa. Data ini didasarkan pada hasil survey dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2023 dimana jumlah ini bertambah cukup signifikan dibandingkan jumlah pengguna pada tahun sebelumnya yang berjumlah 210 juta jiwa (Yati, 2023).

Berdasarkan Datareportal, 60.4% pengguna internet di Indonesia adalah pengguna aktif media sosial seperti Whatsapp, Facebook, Instagram, Tiktok dan sosial media lainnya (Kemp, 2023). Hal ini terjadi karena media sosial berfungsi sebagai sarana komunikasi yang efektif dan efisien.

Berdasarkan informasi yang sudah dipaparkan sebelumnya, dapat terlihat bahwa saat ini kehidupan manusia tidak bisa lepas dari campur tangan teknologi dan internet. Setiap lini kehidupan akan selalu bersinggungan dengan internet mulai dari komunikasi, informasi, pendidikan, hiburan, politik, militer, dan lain sebagainya. Internet sekarang bagaikan kebutuhan pokok yang harus dipenuhi oleh manusia. Hal ini semakin diperkuat sejak adanya pandemi Covid-19 yang memaksa manusia untuk bisa bertahan hidup dan beraktivitas di dalam rumah dengan segala sumber daya yang terbatas. Disinilah awal mula dimana sebagian besar manusia mulai terbiasa dengan kehidupan yang bergantung pada internet. Kini hampir segala hal dapat dilakukan secara online mulai dari bekerja, belajar,

bermain, bersosialisasi, bahkan transaksi jual beli pun kini dapat dilakukan secara online. Otomatis hal ini mempermudah setiap aspek pekerjaan yang dilakukan manusia. Dalam praktiknya, banyak pertukaran data yang terjadi di dalamnya mulai dari pesan, gambar, video dan sebagainya. Seluruh data ini diperlukan sebagai sarana dalam proses pertukaran data guna melakukan proses interaksi maupun transaksi.

Di masa kini, data merupakan sebuah asset berharga yang dapat diolah dan dianalisis menjadi sebuah informasi yang lebih berharga. Bahkan presiden RI Joko Widodo mengatakan bahwa data adalah minyak yang baru yang bahkan lebih bernilai harganya dibandingkan minyak mentah (Nugraheny, 2023). Dengan banyaknya data yang bisa diperoleh saat ini, banyak perusahaan yang mulai mengolah data mereka dan memanfaatkannya untuk meningkatkan efisiensi serta produktifitas perusahaan (Dwitami, 2022). Salah satu cara untuk mengolah dan memanfaatkan data adalah dengan menggunakan pemrosesan bahasa alami atau juga dikenal dengan nama NLP.

Pemrosesan bahasa alami atau NLP (*Natural Language Processing*) adalah ilmu dalam bidang komputer dan bahasa yang berkaitan dengan interaksi antara komputer dan bahasa alami manusia (Katryn, 2020). NLP juga dianggap sebagai salah satu cabang dari kecerdasan buatan yang membantu komputer untuk memahami, menafsirkan dan memanipulasi bahasa manusia. NLP memiliki banyak kegunaan seperti analisis sentimen, deteksi topik, deteksi bahasa, ekstraksi frasa kunci dan kategorisasi dokumen. Saat ini NLP sudah banyak diterapkan dalam berbagai bidang. Contoh yang paling umum adalah pemanfaatan NLP

untuk aplikasi penerjemah, deteksi spam pada *email*, *chatbot*, dan *voice assistant*. Untuk dapat melakukan semua itu, aplikasi NLP membutuhkan banyak data untuk diproses, dianalisis dan dipahami oleh algoritma komputer.

Diantara banyaknya data yang ada, teks merupakan jenis data yang paling sering ditemui dan umum digunakan. Data teks dapat dengan mudah ditemukan di sosial media, forum online, web artikel maupun berita. Namun sama seperti halnya barang tambang lainnya, data teks juga memerlukan tahapan proses terlebih dahulu sebelum dapat diolah menggunakan NLP. Agar bisa mendapatkan hasil yang maksimal, sumber data yang digunakan juga harus relevan (Ranti, 2023) . Meskipun saat ini data teks dapat dengan mudah ditemukan, data teks yang sudah dikumpulkan biasanya masih berbentuk teks yang formatnya tidak terstruktur. Hal ini dapat menyebabkan tidak teridentifikasinya pola pada proses analisis data. Maka teks tersebut harus diubah terlebih dahulu menjadi format yang dapat dipahami oleh komputer.

Salah satu tahap awal dari NLP adalah *text preprocessing*. Tahap ini dilakukan dengan cara mengubah teks jadi bentuk yang lebih terstruktur melalui serangkaian tahap penyesuaian agar dapat lebih mudah diproses oleh komputer. Tahap *text preprocessing* biasanya dilakukan setelah memuat data menggunakan command line Python. Di tahap ini pengguna melakukan penyesuaian teks berdasarkan kebutuhan dan penyesuaian ini dilakukan satu per satu. Setelah itu teks yang sudah dilakukan *preprocessing* dapat diolah lagi untuk dapat menghasilkan informasi baru.

Dalam penelitian ini, penulis berusaha mengembangkan sebuah aplikasi yang dapat dengan mudah digunakan untuk melakukan *text preprocessing* dari data teks yang belum terstruktur dengan menyediakan beberapa fitur *preprocessing* yang dapat dipilih pengguna sesuai dengan kebutuhan sehingga berguna untuk mempermudah pemrosesan bahasa alami.

Berdasarkan uraian latar belakang tersebut, maka dirancanglah sebuah aplikasi berbasis web untuk melakukan *text preprocessing* menggunakan Python dan Flask. Penelitian ini berjudul “Pengembangan Aplikasi Fitur *Preprocessing* Untuk Pemrosesan Bahasa Alami” sebagai objek penelitian dalam menyelesaikan tugas akhir.

1.2. Rumusan Masalah

Berdasarkan pada latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana cara untuk mempermudah tahapan *preprocessing* untuk pemrosesan bahasa alami?
2. Bagaimana cara membuat aplikasi fitur *preprocessing* yang bisa disesuaikan dengan kebutuhan pengguna?

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah :

1. Mempermudah tahapan *preprocessing* untuk pemrosesan bahasa alami.

2. Aplikasi hasil penerapan dapat melakukan fitur *preprocessing* yang bisa disesuaikan dengan kebutuhan pengguna.

1.4. Batasan Masalah

Supaya penelitian ini bisa lebih fokus, maka perlu adanya batasan masalah penelitian. Adapun batasan masalah pada penelitian ini adalah file yang dapat diupload untuk diterapkan fitur *preprocessing* hanya dapat menggunakan file berformat csv dan perancangan aplikasi hanya dibatasi pada beberapa fitur *preprocessing* saja yaitu :

1. *Remove username*
2. *Remove retweet*
3. *Remove hashtag*
4. *Remove url / http*
5. *Remove punctuation*
6. *Remove symbol*
7. *Remove number*
8. *Remove duplicate*
9. *Remove slang*
10. *Remove abbreviation*
11. *Remove elongated character*
12. *Lowercase*
13. *Remove stopword*
14. *Joincase*
15. *Tokenizing*

1.5. Manfaat Penelitian

Hasil dari penelitian ini diharapkan memiliki manfaat sebagai berikut :

1. Membantu penulis dalam memahami serta mengimplementasikan ilmu yang didapat selama perkuliahan untuk memecahkan permasalahan dengan ada pada salah satu bidang ilmu pengetahuan yaitu kecerdasan buatan, khususnya implementasi aplikasi fitur *preprocessing* untuk pemrosesan bahasa alami.
2. Aplikasi yang dihasilkan dapat dimanfaatkan oleh pengguna untuk mempermudah tahap *preprocessing*.
3. Dapat dijadikan sebagai referensi dalam implementasi dan pengembangan aplikasi untuk pemrosesan bahasa alami yang lebih kompleks.

1.6. Keaslian Penelitian

Berdasarkan studi Pustaka yang sudah dilakukan, belum ditemukan adanya penelitian tentang “Pengembangan Aplikasi Fitur *Preprocessing* Untuk Pemrosesan Bahasa Alami”

1.7. Sistematika Penelitian

Sistematika penulisan tugas akhir ini disusun berdasarkan beberapa bab dimana setiap bab membahas mengenai pokok-pokok pembahasan. Adapun sistematika penulisannya adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, keaslian penelitian dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab ini berisi tinjauan pustaka dan landasan teori yang berkaitan dengan penelitian.

BAB III METODE PENGEMBANGAN SYSTEM

Bab ini menjelaskan mengenai metode yang digunakan dalam pengembangan aplikasi.

BAB IV ANALISIS DAN PERANCANGAN SYSTEM

Bab ini berisi tentang analisis kebutuhan system serta langkah-langkah yang dilakukan dalam perancangan system.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan mengenai langkah-langkah implementasi serta tahap pengujian sistem.

BAB VI HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil serta pembahasan dari penelitian.

BAB VII PENUTUP

Bab ini berisikan kesimpulan dan saran untuk penelitian selanjutnya.

BAB VII

PENUTUP

7.1. Kesimpulan

Berdasarkan hasil dari penelitian yang dilakukan terkait dengan pengembangan aplikasi fitur *preprocessing*, maka dapat diambil kesimpulan sebagai berikut :

1. Aplikasi Fitur *Preprocessing* dapat mempermudah pengguna dalam menerapkan fitur *preprocessing*.
2. Aplikasi Fitur *Preprocessing* dapat dengan mudah digunakan untuk melakukan fitur *preprocessing* sesuai dengan kebutuhan pengguna.

7.2. Saran

Terkait hasil dari pengembangan aplikasi fitur *preprocessing* ini tidak terlepas dari adanya kelemahan serta kekurangan. Oleh sebab itu penulis berusaha memberikan saran sebagai pertimbangan untuk pengembangan penelitian selanjutnya, yaitu sebagai berikut :

1. Karena batasan penelitian ini hanya dapat mengimplementasikan beberapa fitur *preprocessing*, aplikasi dapat dikembangkan lagi dengan menambahkan fitur *preprocessing* lain yang lebih kompleks.
2. Karena batasan aplikasi yang hanya dapat mengolah file berekstensi *.csv*, aplikasi dapat dikembangkan lagi dengan menambahkan fitur untuk *formatting* untuk *file* yang *diupload* maupun *file* yang akan *download*.

3. Karena aplikasi ini masih merupakan tahap awal dari alur pemrosesan bahasa alami, maka aplikasi ini dapat dikembangkan lebih jauh lagi menjadi aplikasi pemrosesan bahasa alami yang dapat melakukan pemrosesan data tanpa perlu melakukan koding sama sekali sehingga dapat digunakan oleh siapapun dan kapanpun.
4. Berdasarkan hasil pengujian *beta* non-fungsional oleh pengguna, tampilan aplikasi dapat dibuat lebih menarik lagi agar dapat meningkatkan pengalaman pengguna.

DAFTAR PUSTAKA

- 2023 Developer Survey*. (2023, June 13). Survey.Stackoverflow.Co.
<https://survey.stackoverflow.co/2023/#methodology-general>
- Dicoding Intern. (2021, May 12). *Apa itu UML? Beserta Pengertian dan Contohnya*.
 Dicoding.Com. <https://www.dicoding.com/blog/apa-itu-uml/>
- Dwitami, N. (2022, January 3). *Mengapa Mengolah Data Perusahaan itu Penting?*
 Sunartha.Co.Id. <https://sunartha.co.id/blog/mengapa-mengolah-data-perusahaan-itu-penting/>
- Hasibuan, A. Y. A. (2018). *RANCANG BANGUN APLIKASI PERSEWAAN BUKU BERBASIS WEB MENGGUNAKAN METODE EXTREME PROGRAMMING*.
- Katryn, R. G. (2020, August 28). *Text Preprocessing: Tahap Awal dalam Natural Language Processing (NLP)*. Medium.Com. <https://medium.com/mandiri-engineering/text-preprocessing-tahap-awal-dalam-natural-language-processing-nlp-bc5fbb6606a>
- Kemp, S. (2023, February 9). *DIGITAL 2023: INDONESIA*. Datareportal.Com.
<https://datareportal.com/reports/digital-2023-indonesia>
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Mengenal Flask, Library Machine Learning Python Idaman Developer*. (2021, September 1). Dqlab.Id. <https://dqlab.id/mengenal-flask-library-machine-learning-python-idaman-developer>
- Naseem, U., Razzak, I., Khan, S. K., & Prasad, M. (2021). A Comprehensive Survey on Word Representation Models: From Classical to State-Of-The-Art Word

Representation Language Models. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1). <https://doi.org/10.1145/1122445.1122456>

Noviana, F. (2022). *Sistem Informasi Pemesanan Jasa Fotografi Berbasis Web Di Gulas Media Menggunakan Metode Extreme Programming*.

Nugraheny, D. E. ; S. B. (2023, February 9). *Jokowi: Data adalah “New Oil”, Keamanannya Harus Jadi Perhatian*. Kompas.Com.
<https://nasional.kompas.com/read/2023/02/09/16575621/jokowi-data-adalah-new-oil-keamanannya-harus-jadi-perhatian>

Oktriwina, A. S. (2021, February 2). *NLP: Kecerdasan Buatan yang Bantu Komputer Pahami Bahasa Manusia*. Glints.Com. <https://glints.com/id/lowongan/natural-language-processing-adalah/>

Ranti, S. (2023, February 19). *Pengertian dan Siklus Pemrosesan Data Pada Komputer*. Kompas.Com.
<https://tekno.kompas.com/read/2023/02/19/03000017/pengertian-dan-siklus-pemrosesan-data-pada-komputer-?page=all>

Setyawan, F. E. (2019). *Rancangan Bangun Sistem Informasi Supervisi Akademik Guru Berbasis WEB Dengan Metode Extreme Programming*.

Tahapan Text Preprocessing dalam Teknik Pengolahan Data. (2021, June 17). Dqlab.Id. <https://dqlab.id/tahapan-text-preprocessing-dalam-teknik-pengolahan-data>

Yati, R. (2023, March 8). *Survei APJII: Pengguna Internet di Indonesia Tembus 215 Juta Orang*. Tekno.Bisnis.Com.
<https://teknologi.bisnis.com/read/20230308/101/1635219/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang>



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA