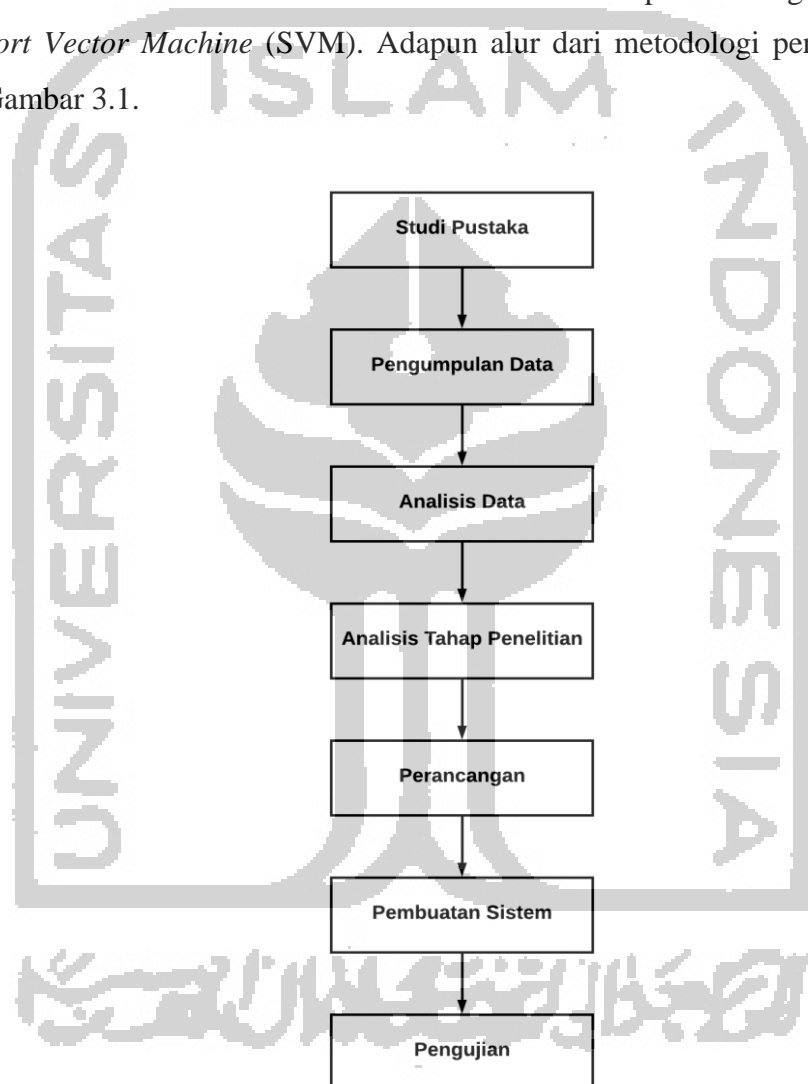


BAB III

METODOLOGI PENELITIAN

Pada bab ini, peneliti akan menjelaskan mengenai metodologi penelitian yang akan dilakukan terkait ekstraksi informasi dan klasifikasi keluhan pasien dengan menggunakan metode *Support Vector Machine* (SVM). Adapun alur dari metodologi penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Alur Metodologi Penelitian

3.1 Studi Pustaka

Studi pustaka yang dilakukan adalah mencari dan mempelajari dari berbagai sumber mengenai ekstraksi informasi pada keluhan pasien. Sumber yang dipelajari berupa jurnal-jurnal, skripsi, tesis, buku, dan referensi dari Internet yang berhubungan dengan sistem penerapan

ekstraksi informasi statistik pada keluhan pasien. Untuk daftar referensi yang digunakan pada penelitian ini dapat dilihat pada daftar pustaka.

3.2 Pengumpulan Data

Dalam penelitian ini, data yang digunakan adalah data primer yang diperoleh dari dr. Rahma, dr. Ari, dr. Dea, dan dr. Alan yang praktik di Klinik Unisia Polifarma UII. Data yang dikumpulkan berupa rekaman suara saat pasien sedang berkonsultasi langsung dengan dokter mengenai keluhan penyakit yang sedang dialami oleh pasien. Hasil dari rekaman suara pasien tersebut kemudian diubah menjadi data teks yang ditulis secara manual. Karakteristik rekaman pasien yang digunakan adalah rekaman yang pasiennya memiliki keluhan sakit, apabila pasien hanya melakukan cek kesehatan maka rekaman pasien tersebut tidak digunakan dalam penelitian ini. Hasil dari pengumpulan data disusun dalam sebuah *file* untuk nantinya diklasifikasi oleh dokter berdasarkan komponen keluhan pasien (keluhan utama, onset, keluhan lain, keterangan, frekuensi serangan, sifat serangan, durasi, lokasi, perjalanan penyakit, riwayat pengobatan sebelumnya, dan akibat gangguan yang timbul).. Selanjutnya, data keluhan pasien yang telah diubah ke dalam bentuk teks tersebut diajukan lagi kepada dokter yang sama untuk dilakukan pengklasifikasian ke dalam komponen keluhan yang telah disebutkan di atas. Selain itu, dalam penelitian ini juga digunakan data sekunder yaitu data yang diperoleh dari *website* <https://www.alodokter.com/>. Hal ini dimaksudkan untuk memperbanyak data. Adapun contoh dari kumpulan teks keluhan pasien dapat dilihat pada Gambar 3.2.

A	B
1	No
2	1 Batuk, pilek, demam sudah dua hari, minum obat sirup, tapi obatnya habis dok.
3	2 Demam dari tadi malam, sama radang dok, kaki sama tangan dingin, terus kejang-kejang karna panas.
4	3 Dari tadi pagi, sakit kepala dok, dari pas bangun tidur biasa aja cuma waktu mulai aktivitas itu udah sakit banget kayak ada beban.
5	Ini sakit terus-menerus dok, kaya kencang banget gitu, di kepala bagian belakang, belum berobat dok, ganggu aktivitas kuliah ini dok.
6	4 Radang tenggorokan dan pilek aja, sejak tadi malam, cuma radangnya hilang timbul gitu sih dok.
7	Belum berobat dok, buat minum sama makan susah buat menelan dok, apalagi kalo kering banget tenggorokan.
8	5 Semenjak dua hari yang lalu tidak enak badan, tenggorokan batuk kering gitu, terus bersin2 sampe pilek sering banget.
9	Terus kalau malem itu kadang bangun badan kayak goyang gitu walaupun guling, terus pusing juga apalagi kalau lihat laptop, uda tapi kadang sama panas juga
10	6 Gatal2 di leher kiri sama kanan, yo satu bulan sudah lama, kalo pas opo sumuk itu keringat, belum diobati dok.
11	Awalnya kecil terus mungkin melebar jadi besar, tidak nyaman ini dok, hilang timbul yo terganggu keluar keringat apa nggak, ini sedikit pusing sih dok.
12	7 Kemarin kan sempet sakit udah periksa di puskesmas, kemaren batuk, serak, terus pusing, sama demam juga, katanya radang.
13	Sudah dikasih obat dan antibiotik nya habis, tapi sekarang batuknya masih, batuk berdahak, terus sekarang tenggorokannya masih sakih sakit terus sekarang ta
14	Kemaren sempet pusing tapi sekarang udah agak mendingan, ke puskesmas udah lima hari yang lalu tapi belum membaik.
15	8 Dada terasa panas dari minggu malam sekitar tiga hari, di apa bagian kiri dada sakit, muncul saat tidur minimal satu kali, hilang timbul gitu, kayak nyeri perut.
16	Ada mual kadang2 dok, udah minum obat tablet beli di apotik sini, aktivitas baik aja sih dok, masih bisa kuliah.
17	9 Batuk, pilek, udah hampir dua minggu, mungkin kemarin itu pas minum air aqua di freezer dingin.
18	Gatal tenggorokan, meriang juga, hanya minum laserin sama biasa jeruk nipis sama itu.
19	10 Keluhannya tiga hari ini saya pusing sama lemes aja, yang cenut2, ya kambuh-kambuhan sih, kalo susah tidur enggak, ada demam juga, kadang panas gitu kan.
20	Kadang nanti apa kayak kemarin hari pertama itu kan panas badannya terus kemarin kan saya pake untuk puasa dia itu agak reda ya.
21	Tadi pagi bangun tidur panas lagi. Tadi ijin kerja juga, jadi tidak bisa kerja, kemarin baru minum vitamin dari poliklinik kantor.
22	11 Asaam lambng dari kemarn siang, sering sampe kantor rasanya mual2.
23	Sakitnya itu di perut kiri ke bawah, terasa sekali saat duduk lama, tidak nyaman di perut .
24	Tadi saya habis makan santan apa itu ada hubungannya ya dok, saya belum berobat sih dok .
25	Sakitnya itu terus-menerus dok .
26	12 Tadi malem tiba2 kuping sebelah kiri kaya ada yang nahan gitu suaranya, berdengung, nggak abis berenang sih, pas diginiin agak sakit dok, dikorek-korek pake c
27	13 Keluhannya cuma batuk kering sama pilek, sudah dua hari.
28	14 Batuk, pilek dua hari, sama demam juga tadi malam dok, belum minum obat dok.
29	15 Gatal di tenggorokan dok, kayaknya habis makan rambutan dok, dari kemarin siang dok.
30	16 Sakit tulang rusuk dok, karna terpeleset, ada migrain juga dok, hilang timbul dari tadi malam.
31	17 Batuk berdahak sama pilek, sudah minum obat sirup dok.
32	18 Cuma batuk, sama agak dada sesak saja dok.
33	19 Batuk kering, pilek, sama tenggorokan gatal, demam juga sudah tiga hari dok. Belum minum obat dok, agak pusing dok.
34	20 Batuk berdahak sama panas dari lima hari dok baru minum obat dari puskesmas itu dok.
35	21 Digigit ini apa, apa semut atau apa, ini sudah dua hari sampe ngilu gini dok, sama gatal. Awalnya kecil terus melebar ini dok.
36	22 Keluhannya pilek, batuk, kadang kadang panas, sakit tenggorokan, dari hari sabtu sore, belum minum obat, tadi tu disengat tawon terus langsung pake fresh car
37	22 kulit kaki saya ini gatal nggak tau iritasi atau apa, dari hari jumat, hilang timbul, belum dikasih obat dok, apa mungkin karena kost nggak ditingin habis diting
38	pilek, terus batuk, terus tenggorokan gatal, terus tadi malam lemas, belum minum obat, nggak merokok, tapi kalo kena orang yang merokok sakit.
39	23 makan tidak nyaman, bersin-bersin terus kena ac gitu meler juga nggak tau alergi karna apa.

Gambar 3.2 Teks Keluhan Pasien

3.3 Analisis Data

Data yang telah dikumpulkan, lalu dilakukan analisis mengenai struktur pada data yang akan dikelompokkan berdasarkan komponennya. Adapun analisis yang didapat pada teks klinis adalah sebagai berikut:

- Singkatan Kata
Contoh: yg, dr, td, skt.
- Pengulangan Kata
Contoh: mual2, paru2, makan2.
- Kesalahan Penulisan Kata.
Contoh: serting, karna, yng, sudah, belm.
- Stopword Kata
Contoh: dari, sampai, ada, karena.

Gambar 3.3 merupakan salah satu contoh teks keluhan pasien dari dokter yang akan dianalisis untuk proses selanjutnya.

Sakit tulang rusuk dok, karna terpeleset, ada migrain juga dok, hilang timbul dr td malam. Sma mual2 dok.

Gambar 3.3 Contoh Teks Keluhan Pasien Dari Dokter

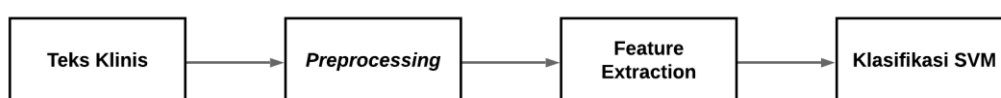
Tabel 3.1 merupakan salah satu contoh dari data teks klinis yang telah dianalisis dan dipetakan ke dalam komponen keluhan pasien yang sesuai.

Tabel 3.1 Contoh Pemetaan Teks Klinis

Riwayat Penyakit Sekarang	Kalimat ke-1	Kalimat ke-2
Keluhan Utama	sakit tulang rusuk	Mual-mual
Onset	tadi malam	
Keluhan Lain	migrain	
Keterangan		
Frekuensi Serangan		
Sifat Serangan	hilang timbul	
Durasi		
Lokasi		
Perjalanan Penyakit	terpeleset	
Riwayat Pengobatan Sebelumnya		
Akibat Gangguan yang Timbul		

3.4 Analisis Tahap Penelitian

Analisis tahap penelitian merupakan sebuah gambaran umum dari tahap yang akan dilakukan dalam membangun sistem pada penelitian ini agar lebih terstruktur. Adapun alur dari tahap penelitian dapat dilihat pada Gambar 3.4



Gambar 3.4 Alur Tahap Penelitian

3.4.1 Teks Klinis

Teks klinis merupakan informasi berupa data teks dari keluhan pasien yang akan dimasukkan ke dalam sistem sebagai data masukan. Teks klinis dapat berisi satu kalimat atau lebih di mana satu kalimat dapat terdiri dari satu frasa/klausa atau lebih. Contoh teks keluhan pasien dari hasil rekaman pasien yang diambil di Klinik Unisia Polifarma dapat dilihat pada Gambar 3.5. dan contoh teks keluhan pasien yang diambil dari *website* alodoktercom dapat dilihat pada Gambar 3.6.

Selamat pagi dok???? nama saya sukron, umur 20 thn. Saya asaam lambng dr kemarn siang, serring sendawa sampe kantor rasanya mual2. Sakitnya itu di perut kiri ke bawah, terasa sekali saat duduk lama, tidak nyaman di perut. Tadi saya habis makan santan, apa itu ada hubungannya ya dok? saya belum berobat sih dok. Sakitnya itu terus-menerus dok

Gambar 3.5 Contoh Teks Keluhan Pasien



Tian
Anggota

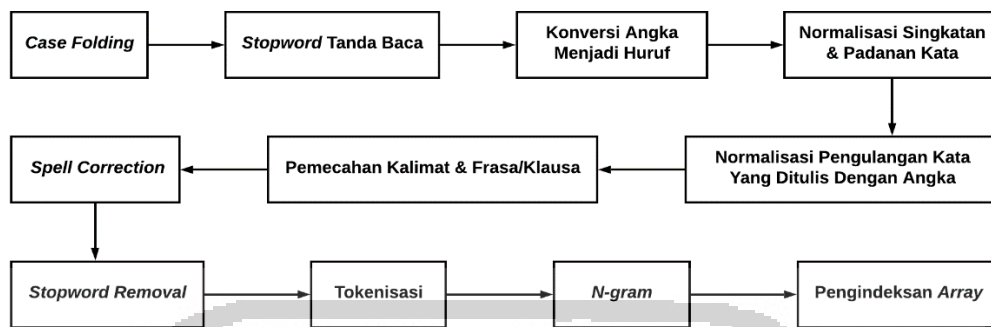
19 April 2019, 15:31

Selamat sore dok maaf saya sudah dua hari ini demam dan semalam mulai muncul bintik merah seperti melepuh yang gatal sekali ... Dan sekarang semakin banyak sampai ke muka .. Umur saya 26

Gambar 3.6 Contoh Teks Alodokter

3.4.2 Preprocessing

Preprocessing merupakan sebuah tahap awal atau pemrosesan yang digunakan untuk mengelola data pada teks keluhan pasien sebelum memasuki proses tahap utama yaitu klasifikasi SVM. Tahap *preprocessing* pada penelitian ini terdiri dari proses *case folding*, *stopword* tanda baca, konversi angka menjadi huruf, normalisasi singkatan dan padanan kata, normalisasi pengulangan kata yang ditulis dengan angka, pemecahan kalimat dan frasa/klausa, *spell correction*, *stopword removal*, tokenisasi, *n-gram*, dan pengindeksan *array*. Adapun tahap *preprocessing* dapat dilihat pada Gambar 3.7.



Gambar 3.7 Tahap *Preprocessing*

Adapun penjelasan dari urutan tahapan *preprocessing* adalah sebagai berikut:

a. *Case Folding*

Case folding merupakan sebuah proses yang mengubah data teks keluhan pasien menjadi huruf kecil semua. Pada tahap ini, setelah tanda baca koma (,) dan titik (.) diberi spasi agar dapat memudahkan proses selanjutnya. Contoh penerapan pada *case folding* dapat dilihat pada Tabel 3.2.

Tabel 3.2 Contoh Proses *Case Folding*

Teks Klinis	<i>Case Folding</i>
Selamat pagi dok???? nama saya sukron, umur 20 thn. Saya asaam lambng dr kemarn siang, serring sendawa sampe kantor rasanya mual2. Sakitnya itu di perut kiri ke bawah, terasa sekali saat duduk lama, tidak nyaman di perut. Tadi saya habis makan santan, apa itu ada hubungannya ya dok? saya belum berobat sih dok. Sakitnya itu terus-menerus dok	selamat pagi dok???? nama saya sukron , umur 20 thn . saya asaam lambng dr kemarn siang , serring sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok? saya belum berobat sih dok . sakitnya itu terus-menerus dok

b. *Stopword* Tanda Baca

Stopword tanda baca merupakan proses untuk menghilangkan tanda baca selain titik (.), koma (,), dan hubung (-). Hal ini dilakukan dikarenakan terdapat keluhan pasien (data alodokter) yang menggunakan tanda tanya (?) lebih dari satu. Sebagai contoh adalah

“selamat pagi dok???? nama saya sukron”. Adapun penerapan *stopword* tanda baca dapat dilihat pada Tabel 3.3.

Tabel 3.3 Contoh Proses *Stopword* Tanda Baca

<i>Case Folding</i>	<i>Stopword</i> Tanda Baca
selamat pagi dok???? nama saya sukron , umur 20 thn . saya asaam lambng dr kemarn siang , serting sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok? saya belum berobat sih dok . sakitnya itu terus-menerus dok	selamat pagi dok nama saya sukron , umur 20 thn . saya asaam lambng dr kemarn siang , serting sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok saya belum berobat sih dok . sakitnya itu terus-menerus dok

c. Konversi Angka Menjadi Huruf

Konversi angka menjadi huruf merupakan proses yang mengubah angka menjadi huruf untuk memudahkan pencocokan kata pada kamus. Contohnya adalah “20” menjadi “dua puluh”. Contoh penerapan tahap konversi angka menjadi huruf dapat dilihat pada Tabel 3.4.

Tabel 3.4 Contoh Proses Konversi Angka Menjadi Huruf

<i>Stopword</i> Tanda Baca	Konversi Angka Menjadi Huruf
selamat pagi dok nama saya sukron , umur 20 thn . saya asaam lambng dr kemarn siang , serting sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok saya belum berobat sih dok . sakitnya itu terus-menerus dok	selamat pagi dok nama saya sukron , umur dua puluh thn . saya asaam lambng dr kemarn siang , serting sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok saya belum berobat sih dok . sakitnya itu terus-menerus dok

d. Normalisasi Singkatan & Padanan Kata

Normalisasi singkatan dan padanan kata merupakan sebuah proses yang menormalisasikan kata dari kata yang disingkat dan padanan dari kata yang memiliki arti sama agar memudahkan dalam proses selanjutnya. Contohnya adalah kata “dr” diubah menjadi “dari”, “dok”, menjadi “dokter”, dan “mencret” menjadi “diare”. Contoh penerapan tahap normalisasi singkatan dan padanan kata dapat dilihat pada Tabel 3.5.

Tabel 3.5 Contoh Proses Normalisasi Singkatan & Padanan Kata

Konversi Angka Menjadi Huruf	Normalisasi Singkatan & Padanan Kata
selamat pagi dok nama saya sukron , umur dua puluh thn . saya asaam lambng dr kemarn siang , serring sendawa sampe kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya ya dok saya belum berobat sih dok . sakitnya itu terus-menerus dok	selamat pagi dokter nama saya sukron , umur dua puluh tahun . saya asaam lambng dari kemarn siang , serring sendawa sampai kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya iya dokter saya belum berobat sih dokter . sakitnya itu terus-menerus dokter

e. Normalisasi Pengulangan Kata Yang Ditulis Dengan Angka

Normalisasi pengulangan kata yang ditulis dengan angka merupakan sebuah proses untuk menormalisasikan kata singkatan yang ditulis dengan menggunakan angka 2. Contohnya adalah kata “mual2” menjadi “mual-mual”. Contoh penerapan tahap normalisasi pengulangan kata yang ditulis dengan angka dapat dilihat pada Tabel 3.6.

Tabel 3.6 Contoh Proses Normalisasi Pengulangan Kata Yang Ditulis Dengan Angka

Normalisasi Singkatan & Padanan Kata	Normalisasi Pengulangan Kata Yang Ditulis Dengan Angka
selamat pagi dokter nama saya sukron , umur dua puluh tahun	selamat pagi dokter nama saya sukron , umur dua puluh tahun . saya asaam lambng dari

<p>. saya asam lambng dari kemarn siang , serting sendawa sampai kantor rasanya mual2 . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya iya dokter saya belum berobat sih dokter . sakitnya itu terus-menerus dokter</p>	<p>kemarn siang , serting sendawa sampai kantor rasanya mual-mual . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya iya dokter saya belum berobat sih dokter . sakitnya itu terus-menerus dokter</p>
---	---

f. Pemecahan Kalimat & Frasa/Klausa

Pemecahan kalimat dan frasa/klausa merupakan sebuah proses yang memisahkan antar kalimat berdasarkan tanda baca titik (.) dan memisahkan frasa/klausa berdasarkan tanda baca koma (,) untuk memudahkan proses selanjutnya. Contoh penerapan tahap pemecahan kalimat dan frasa/klausa dapat dilihat pada Tabel 3.7.

Tabel 3.7 Contoh Proses Pemecahan Kalimat & Frasa/Klausa

Normalisasi Pengulangan Kata Yang Ditulis Dengan Angka	Pemecahan Kalimat & Frasa/Klausa
<p>selamat pagi dokter nama saya sukron , umur dua puluh tahun . saya asam lambng dari kemarn siang , serting sendawa sampai kantor rasanya mual-mual . sakitnya itu di perut kiri ke bawah , terasa sekali saat duduk lama , tidak nyaman di perut . tadi saya habis makan santan , apa itu ada hubungannya iya dokter saya belum berobat sih dokter . sakitnya itu terus-menerus dokter</p>	<p>[['selamat pagi dokter nama saya sukron', 'umur dua puluh tahun', ['saya asam lambng dari kemarn siang', 'serting sendawa sampai kantor rasanya mual-mual'], ['sakitnya itu di perut kiri ke bawah', 'terasa sekali saat duduk lama', 'tidak nyaman di perut'], ['tadi saya habis makan santan', 'apa itu ada hubungannya iya dokter saya belum berobat sih dokter'], ['sakitnya itu terus-menerus dokter']]]</p>

g. *Spell Correction*

Spell correction merupakan sebuah proses yang mengubah atau memeriksa kesalahan pada penulisan kata agar menjadi kata yang benar. Pada tahap ini, diperlukan proses

tokenisasi yang ditujukan untuk memisahkan kata agar proses pada *spell correction* dapat dilakukan. Contohnya adalah kata “asaam” menjadi “asam” dan “serting” menjadi “sering”. Contoh penerapan tahap *spell correction* dapat dilihat pada Tabel 3.8.

Tabel 3.8 Contoh Proses *Spell Correction*

Pemecahan Kalimat & Frasa/Klausa	<i>Spell Correction</i>
[[selamat pagi dokter nama saya sukron', 'umur dua puluh tahun'], ['saya asaam lambng dari kemarn siang', 'serting sendawa sampai kantor rasanya mual-mual'], ['sakitnya itu di perut kiri ke bawah', 'terasa sekali saat duduk lama', 'tidak nyaman di perut'], ['tadi saya habis makan santan', 'apa itu ada hubungannya iya dokter saya belum berobat sih dokter'], ['sakitnya itu terus-menerus dokter']]	[[selamat pagi dokter nama saya sukron', 'umur dua puluh tahun'], ['saya asam lambung dari kemarin siang', 'sering sendawa sampai kantor rasanya mual-mual'], ['sakitnya itu di perut kiri ke bawah', 'terasa sekali saat duduk lama', 'tidak nyaman di perut'], ['tadi saya habis makan santan', 'apa itu ada hubungannya iya dokter saya belum berobat sih dokter'], ['sakitnya itu terus-menerus dokter']]

h. *Stopword Removal*

Stopword removal merupakan sebuah proses untuk menghilangkan kata-kata yang tidak diperlukan atau tidak memiliki makna berdasarkan teks keluhan pasien. Adapun daftar kata-kata yang dihilangkan dimasukkan dalam daftar *stopword*. Daftar sebagian kata dari *stopword* dapat dilihat pada Tabel 3.9.

Tabel 3.9 Daftar Sebagian *Stopword*

ada	aku	dari	sekali	harus
hampir	dan	jika	juga	kalau
karena	lagi	sering	sampai	di

Berikut ini merupakan contoh teks keluhan pasien yang telah diproses pada tahap *stop removing*. Hasil dari proses *stopword removal* dapat dilihat pada Tabel 3.10.

Tabel 3.10 Contoh Proses *Stopword Removal*

<i>Spell Correction</i>	<i>Stopword Removal</i>
<p>['selamat pagi dokter nama saya sukron', 'umur dua puluh tahun'], ['saya asam lambung dari kemarin siang', 'sering sendawa sampai kantor rasanya mual-mual'], ['sakitnya itu di perut kiri ke bawah', 'terasa sekali saat duduk lama', 'tidak nyaman di perut'], ['tadi saya habis makan santan', 'apa itu ada hubungannya iya dokter saya belum berobat sih dokter'], ['sakitnya itu terus-menerus dokter']]</p>	<p>['', ''], ['asam lambung kemarin siang', 'sering sendawa mual-mual'], ['perut kiri bawah', 'saat duduk lama', 'tidak nyaman perut'], ['makan santan', 'belum berobat'], ['terus-menerus']]</p>

i. Tokenisasi

Tokenisasi merupakan sebuah proses untuk memisahkan antar kata atau yang disebut token, agar memudahkan proses pengelompokan kata pada proses selanjutnya. Contoh penerapan pada tokenisasi dapat dilihat pada Tabel 3.11.

Tabel 3.11 Contoh Proses Tokenisasi

<i>Stopword Removal</i>	Tokenisasi
<p>['', ''], ['asam lambung kemarin siang', 'sering sendawa mual-mual'], ['perut kiri bawah', 'saat duduk lama', 'tidak nyaman perut'], ['makan santan', 'belum berobat'], ['terus-menerus']]</p>	<p>[[], []], ['asam', 'lambung', 'kemarin', 'siang'], ['sering', 'sendawa', 'mual-mual'], ['perut', 'kiri', 'bawah'], ['saat', 'duduk', 'lama'], ['tidak', 'nyaman', 'perut'], ['makan', 'santan'], ['belum', 'berobat'], ['terus-menerus']]</p>

j. *N-Gram*

N-gram merupakan sebuah proses pengelompokan kata menjadi satu, dua, tiga, atau *n*-kata. Adapun *n-gram* yang digunakan dalam penelitian ini adalah *tri-gram*, *bi-gram*, dan *uni-gram*. Contoh penerapan tahap *n-gram* dapat dilihat pada Tabel 3.12.

Tabel 3.12 Contoh Proses N-Gram

Tokenisasi	N-Gram
[[], [], ['asam', 'lambung', 'kemarin', 'siang'], ['sering', 'sendawa', 'mual-mual'], ['perut', kiri', 'bawah'], ['saat', 'duduk', 'lama'], ['tidak', 'nyaman', 'perut'], ['makan', 'santan'], ['belum', 'berobat'], ['terus-menerus']]	['asam lambung', 'kemarin siang', 'sering sendawa', 'mual-mual', 'perut kiri bawah', saat duduk lama', 'tidak nyaman perut', makan santan', 'belum berobat', 'terus- menerus']

k. Pengindeksan *Array*

Pengindeksan *array* merupakan proses untuk memberikan indeks pada *array* agar dapat menentukan kata/frasa/klausa berasal dari kalimat ke berapa. Contoh penerapan tahap pengindeksan *array* dapat dilihat pada Tabel 3.13.

Tabel 3.13 Contoh Proses Pengindeksan *Array*

N-Gram	Pengindeksan Array
['asam lambung', 'kemarin siang', 'sering sendawa', 'mual-mual', 'perut kiri bawah', saat duduk lama', 'tidak nyaman perut', makan santan', 'belum berobat', 'terus- menerus']	[['asam lambung', 1], ['kemarin siang', 1], ['sering sendawa', 1], ['mual-mual', 1], ['perut kiri bawah', 2], ['saat duduk lama', 2], ['tidak nyaman perut', 2], ['makan santan', 3], ['belum berobat', 3], ['terus-menerus', 4]]

Pada kalimat tersebut terdapat lima kalimat atau (kalimat ke-0, ke-1, ke-2, ke-3, ke-4). Akan tetapi, pada kalimat ke-0 telah dihapus pada saat melewati proses *stopword removal*. Kalimat tersebut dihapus karena termasuk dalam daftar kata *stopword*, sehingga hasil pada Tabel 3.13 *array*-nya dimulai pada kalimat ke-1. Setelah dilakukan proses ini, kata/frasa/klausa tersebut dibagi menjadi 2 kelompok, kata/frasa/klausa yang tidak termasuk dalam keluhan (keluhan utama dan keluhan lain) diklasifikasi menggunakan SVM. Sedangkan kata keluhan, diklasifikasi berdasarkan kamus “klasifikasi.txt” dan kata tersebut dibedakan lagi mana yang keluhan utama dan keluhan lain. Kata/frasa/klausa yang akan diklasifikasi menggunakan SVM harus dilakukan proses *feature extraction* terlebih dahulu.

3.4.3 Feature Extraction

Metode yang digunakan untuk klasifikasi teks keluhan pasien pada sistem ini menggunakan TF-IDF. TF-IDF (*Term Frequency – Inverse Document Frequency*) merupakan metode atau algoritma yang akan digunakan untuk menghitung bobot dari teks keluhan pasien. Contoh data yang digunakan dapat dilihat pada Tabel 3.14.

Tabel 3.14 Data yang digunakan

d1	perut kiri bawah
d2	saat duduk lama
d3	tidak nyaman perut
d4	makan santan
d5	terus-menerus

Nilai *term frequency* diberi bobot 1 pada masing-masing kata. Hasil dari *term frequency* dapat dilihat pada Tabel 3.15 dan hasil dari perhitungan *tf-idf* dalam sistem dapat dilihat pada Gambar 3.8.

Tabel 3.15 *Term Frequency*

	d1	d2	d3	d4	d5
perut	1	0	1	0	0
kiri	1	0	0	0	0
bawah	1	0	0	0	0
saat	0	1	0	0	0
duduk	0	1	0	0	0
lama	0	1	0	0	0
tidak	0	0	1	0	0
nyaman	0	0	1	0	0
makan	0	0	0	1	0
santan	0	0	0	1	0
terus	0	0	0	0	1
menerus	0	0	0	0	1

```

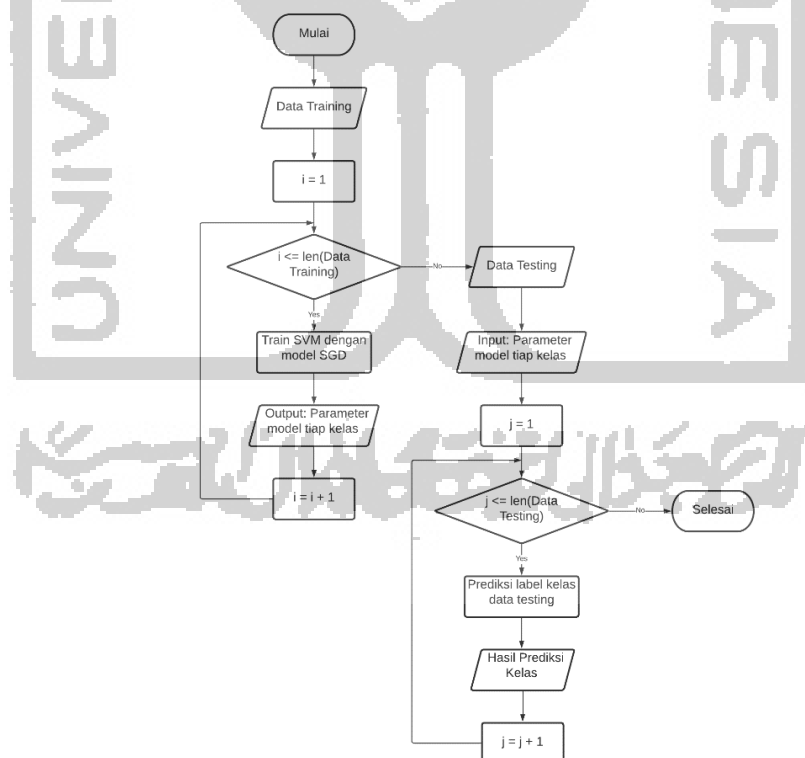
X_tes_tfidf (0, 269) 0.5870152630615575
(0, 170) 0.5467249996447937
(0, 27) 0.5970802757554219
(1, 288) 0.5234743191143233
(1, 184) 0.6024842891012991
(1, 102) 0.6024842891012991
(2, 333) 0.5172463280521703
(2, 269) 0.555364195027552
(2, 242) 0.6511734384923465
(3, 292) 0.7864243930368463
(3, 204) 0.6176865499884451
(4, 330) 0.6680942246285189
(4, 216) 0.7440766808723536

```

Gambar 3.8 Nilai *tf-idf*

3.4.4 Klasifikasi SVM

Metode yang digunakan untuk klasifikasi teks keluhan pasien pada sistem ini menggunakan proses pembelajaran dari SVM. Setelah teks keluhan pasien tersebut diberi bobot nilai dengan *tf-idf*, maka selanjutnya dilakukan proses untuk mengklasifikasikan teks keluhan pasien tersebut ke dalam sembilan kelas Riwayat Penyakit Sekarang (RPS). *Flowchart* untuk klasifikasi SVM dapat dilihat pada Gambar 3.9.

Gambar 3.9 *Flowchart* Klasifikasi SVM

Setiap data *training* dilatih dengan model SGD dan dihitung dengan persamaan 3.1 untuk mendapatkan nilai parameter model dari tiap kelas. SGD (*Stochastic Gradient Descent*) adalah pendekatan sederhana yang efisien untuk pengklasifikasi linear seperti SVM dan mendukung klasifikasi multi-kelas dengan *one vs all*. Parameter model dapat diakses melalui `coef_` dan `intercept_`. `Coef_` adalah array bentuk dua dimensi `[n_classes, n_features]` dan `intercept_` adalah array bentuk satu dimensi `[n_classes]`.

$$w \leftarrow w - \eta \left(\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial L(w^T x_i + b, y_i)}{\partial w} \right) \quad (3.1)$$

Keterangan :

W : parameter model bobot (`coef_`)

η : ukuran parameter model

b : parameter model (`intercept_`)

L : fungsi untuk model fit

R : regularisasi *term* untuk kompleksitas model

y_i : target ke- i

x_i : data ke- i

Setiap data *testing* dimasukkan nilai parameter model dari tiap kelas dan diprediksi dengan nilai maksimum untuk mendapatkan hasil prediksi dari setiap kelas. Hasil penerapan dari klasifikasi SVM dapat dilihat pada Gambar 3.10.

```
Prediksi : ['lokasi' 'keterangan' 'akibatGangguan' 'perjalananPenyakit' 'sifat']
['perut kiri bawah', 1] lokasi
['saat duduk lama', 1] keterangan
['tidak nyaman perut', 1] akibatGangguan
['makan santan', 2] perjalananPenyakit
['terus-menerus', 3] sifat
```

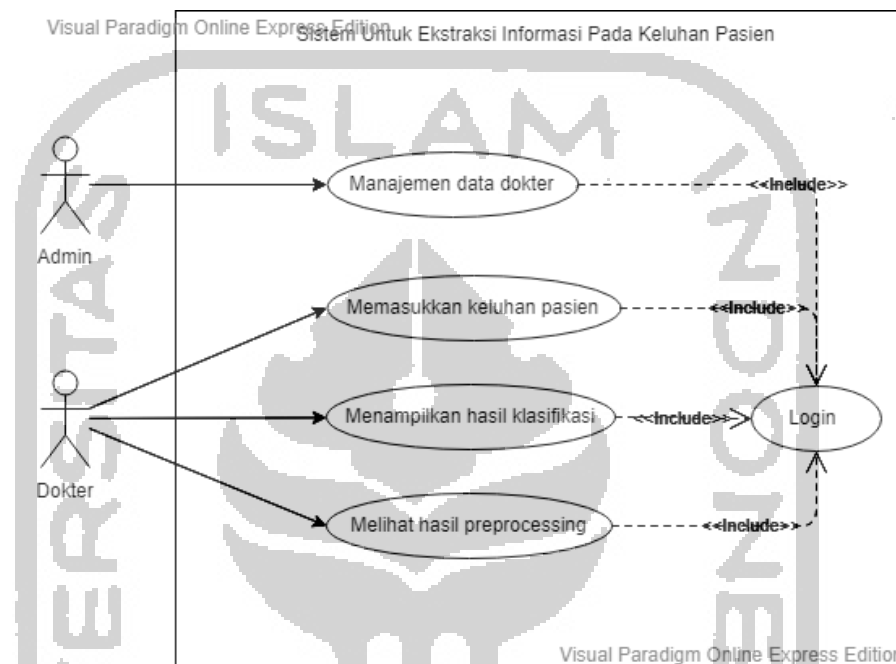
Gambar 3.10 Hasil Penerapan Klasifikasi SVM

3.5 Perancangan

Perancangan sistem dilakukan untuk memudahkan tahap pada implementasi sistem berdasarkan hasil analisis yang dilakukan sebelumnya. Perancangan ini terdiri dari *use case diagram* dan perancangan *interface*.

3.5.1 Use Case Diagram

Sistem untuk ekstraksi informasi pada keluhan pasien menggunakan metode statistik memiliki dua aktor yaitu admin dan dokter. Aksi yang dapat dilakukan oleh masing-masing aktor akan dijelaskan melalui *use case diagram* pada Gambar 3.11.



Gambar 3.11 Use Case Diagram

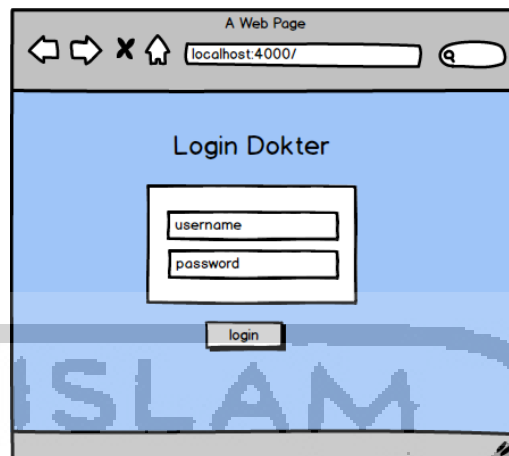
Pada Gambar 3.11, aktor admin dapat melakukan manajemen data dokter meliputi tambah, ubah, dan hapus. Sedangkan aktor dokter dapat memasukkan keluhan pasien, menampilkan hasil klasifikasi, dan melihat hasil *preprocessing*.

3.5.2 Perancangan Interface

Perancangan *interface* merupakan gambaran desain tampilan dari sistem yang akan dibangun pada web. Perancangan *interface* terdiri dari tampilan admin dan dokter. Adapun perancangan *interface* pada sistem tersebut dapat dilihat di bawah ini.

a. Halaman *Login* Dokter

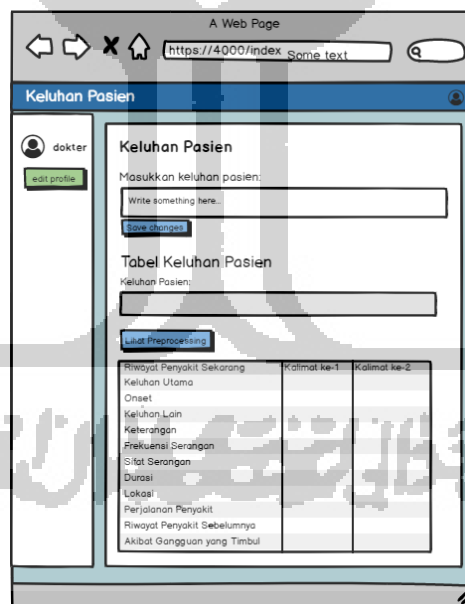
Pada halaman *login* dokter, aktor dapat memasukkan *username* dan *password* untuk masuk ke dalam sistem. Adapun rancangan *interface* pada halaman *login* dokter dapat dilihat pada Gambar 3.12.



Gambar 3.12 Rancangan Halaman *Login Dokter*

b. Halaman Dokter

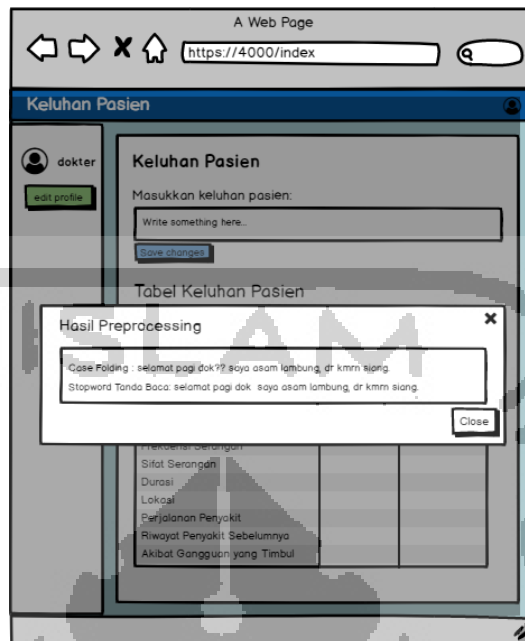
Pada halaman ini, dokter dapat memasukkan keluhan pasien dan melihat hasil klasifikasi dari masukan tersebut pada tabel keluhan pasien. Adapun rancangan *interface* pada halaman dokter dapat dilihat pada Gambar 3.13.



Gambar 3.13 Rancangan Halaman Dokter

c. Halaman Hasil *Preprocessing*

Pada halaman ini menampilkan hasil dari *preprocessing* apabila aktor memilih tombol lihat *preprocessing*. Adapun rancangan *interface* dari halaman hasil *preprocessing* dapat dilihat pada Gambar 3.14.



Gambar 3.14 Rancangan Halaman Hasil *Preprocessing*

d. Halaman *Login Admin*

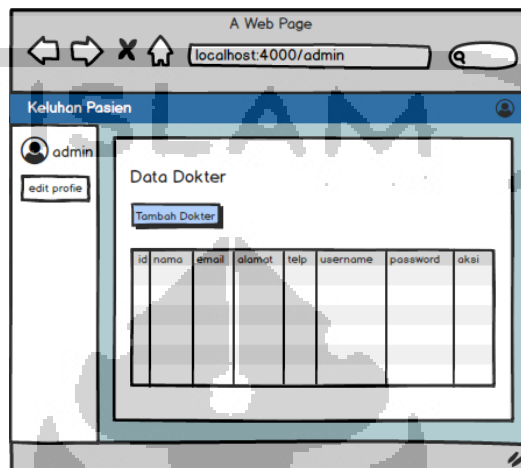
Pada halaman *login admin*, aktor dapat memasukkan *username* dan *password* untuk masuk ke dalam sistem. Adapun rancangan *interface* dari halaman *login admin* dapat dilihat pada Gambar 3.15.



Gambar 3.15 Rancangan Halaman *Login Admin*

e. Halaman Admin

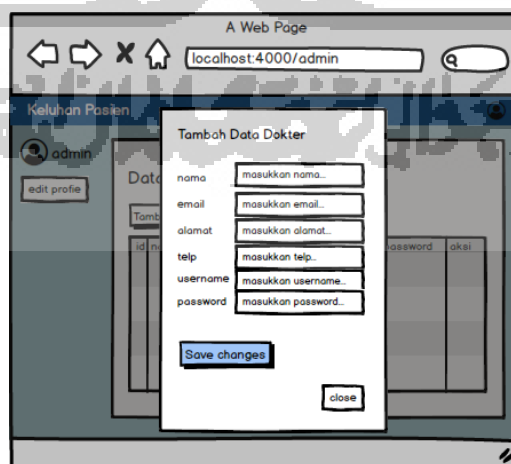
Pada halaman ini, admin dapat melihat data dokter yang ditampilkan dalam bentuk tabel. Admin juga dapat menambah, mengubah, dan menghapus data dokter. Adapun rancangan *interface* pada halaman admin dapat dilihat pada Gambar 3.16.



Gambar 3.16 Rancangan Halaman Admin

f. Halaman Tambah Dokter

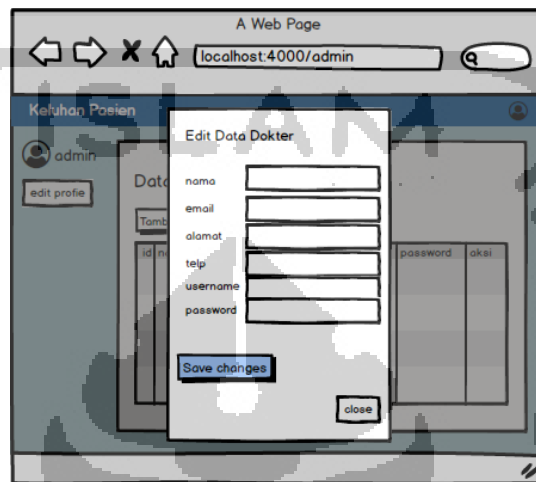
Pada halaman tambah dokter, admin dapat menambahkan data dokter baru dengan menambahkan informasi seperti: nama, email, alamat, nomor telpon, *username* dan *password*. Adapun rancangan *interface* pada halaman tambah dokter dapat dilihat pada Gambar 3.17.



Gambar 3.17 Rancangan Halaman Tambah Dokter

g. Halaman *Edit Dokter*

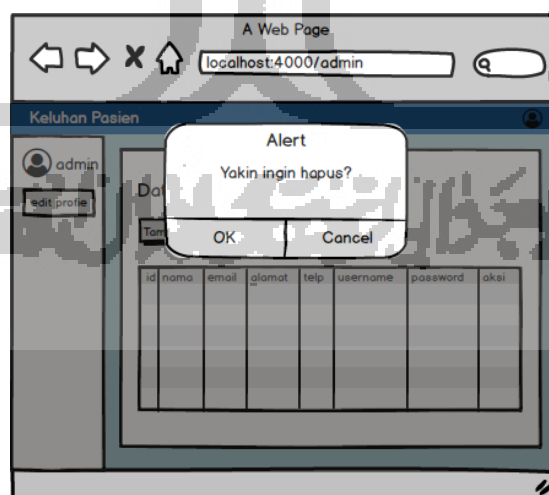
Pada halaman *edit dokter*, admin dapat mengubah data dokter seperti: nama, *email*, alamat, nomor telpon, *username* dan *password*. Adapun rancangan *interface* pada halaman *edit dokter* dapat dilihat pada Gambar 3.18.



Gambar 3.18 Rancangan Halaman *Edit Dokter*

h. Halaman Hapus Dokter

Pada halaman hapus dokter, admin dapat menghapus data dokter. Adapun rancangan *interface* pada halaman hapus dokter dapat dilihat pada Gambar 3.19.



Gambar 3.19 Rancangan Halaman Hapus Dokter

3.6 Pembuatan Sistem

Sistem ini dibangun menggunakan bahasa pemrograman Python dan dijumpatani dengan *browser*. Untuk semua proses pada keluhan data pasien dikelola menggunakan *software* JetBrains PyCharm, di mana proses untuk Python dan *html* dipisah. Untuk memanggil fungsi pada Python agar dapat ditampilkan ke *browser* menggunakan *web framework* Flask dan Jinja.

3.7 Pengujian

Pengujian ini nantinya akan dilakukan di Klinik Unisia Polifarma UII. Nantinya akan dilakukan perbandingan dari pengklasifikasian yang dilakukan oleh dokter dengan hasil klasifikasi yang dilakukan oleh sistem, sehingga dapat diperoleh tabel perbandingan hasil klasifikasi. Sedangkan pengujian untuk metode SVM menggunakan pengujian *confusion matrix*.

Confusion matrix adalah matriks yang sering digunakan untuk mengukur kinerja suatu algoritma klasifikasi. *Confusion matrix* berisi informasi tentang klasifikasi aktual dan prediksi dilakukan dengan klasifikasi sistem yang dievaluasi menggunakan data dalam matriks (Santra & Christy, 2012). *Confusion matrix* dapat dilihat pada Tabel 3.16.

Tabel 3.16 *Confusion Matrix*

	<i>Predict : Positive</i>	<i>Predicted : Negative</i>
<i>Actual : Positive</i>	TP	FP
<i>Actual : Negative</i>	FN	TN

Dari *confusion matrix* tersebut, dapat dihitung nilai *accuracy*, *precision*, *recall*, dan *f1-score* dengan persamaan sebagai berikut:

$$Accuracy = \frac{TP}{TP + FP + TN + FN} \times 100\% \quad (3.2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3.3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3.4)$$

$$F1\text{-score} = 2 \times \frac{(\textit{Precision} + \textit{Recall})}{(\textit{Precision} + \textit{Recall})} \times 100\% \quad (3.5)$$

Keterangan:

TP (*True Positive*) : data positif yang diprediksi benar sebagai data positif

TN (*True Negative*) : data negatif yang diprediksi benar sebagai data data negatif

FP (*False Positive*) : data positif yang diprediksi salah sebagai data negatif

FN (*False Negative*): data negatif yang diprediksi salah sebagai data positif

Perbedaan antara *macro average*, *micro average*, dan *weighted average* adalah *macro average* menghitung matriks secara bebas untuk setiap kelas kemudian mengambil rata-ratanya dan *micro average* menghitung matriks rata-rata dengan mengumpulkan kontribusi dari semua kelas. Sedangkan *weigthed average*, akan menghitung rata-rata dengan memperhitungkan bobot pada setiap datanya.

