

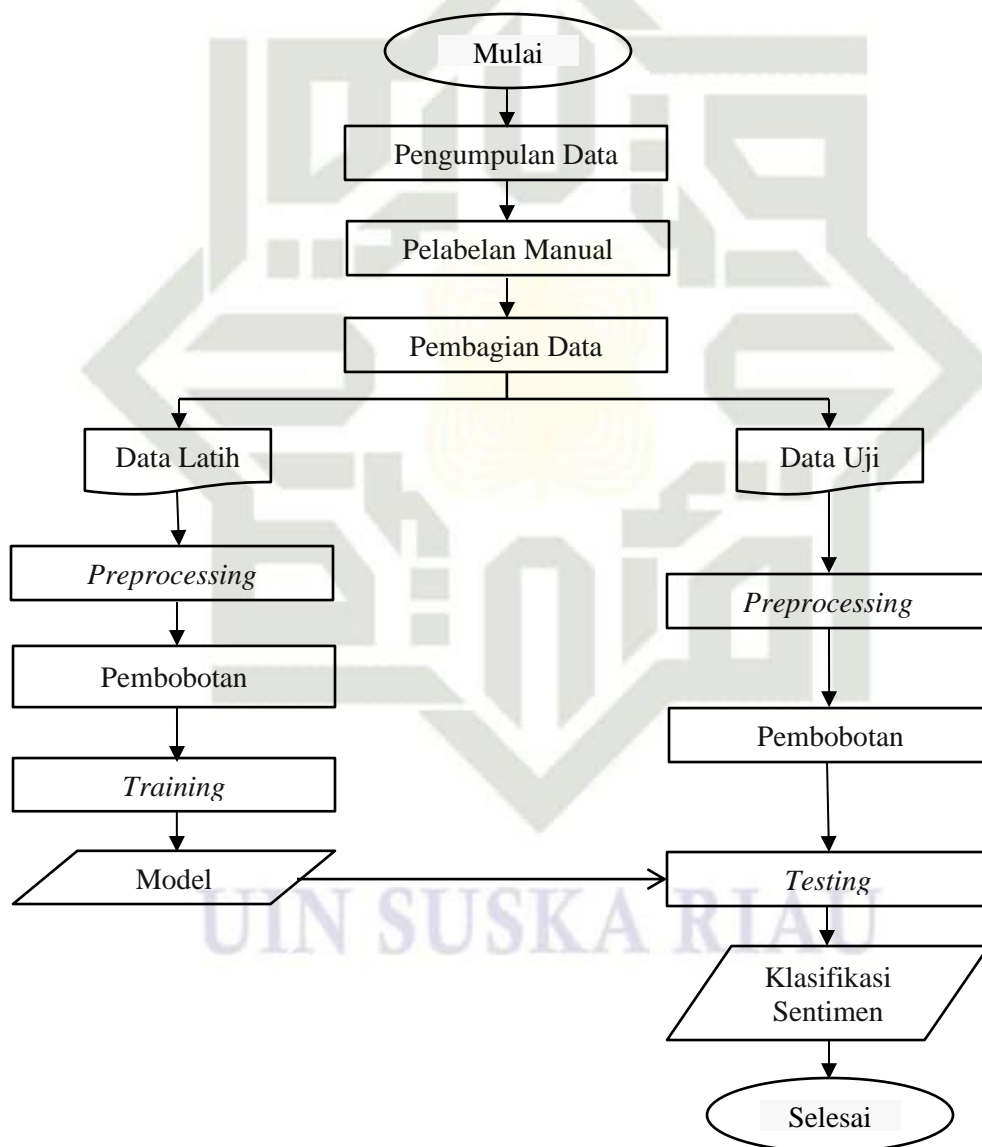
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB IV

ANALISA DAN PERANCANGAN

Tahapan analisa dan perancangan merupakan tahapan penting pada penelitian yang akan dilakukan. Pada bab ini akan dilakukan pembahasan mengenai analisa sumber data, proses pelabelan manual, praproses data, proses pembobotan dan bagaimana proses klasifikasi sentimen menggunakan metode *support vector machine*.



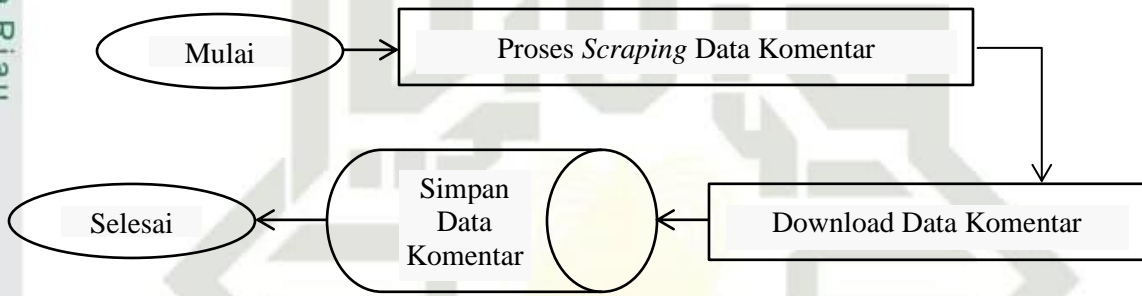
Gambar 4.1 Proses Klasifikasi *Support Vector Machine*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.1 Analisa

Penelitian ini akan melakukan bagaimana cara mengetahui komentar yang berbentuk sentimen terhadap Ustad Abdul Somad. Komentar-komentar tersebut nantinya akan dianalisa untuk menentukan apakah komentar tersebut masuk kedalam kategori sentimen negatif atau positif menggunakan metode *support vector machine*. Proses pengambilan data komentar dilakukan dengan cara *scraping* menggunakan tools *Spyder*. Proses *scraping* data komentar pada video *channel official* Ustad Abdul Somad dapat dilihat pada *flowchart* dibawah.



Gambar 4.2 Proses *Scraping* Data Komentar

4.1.1 Analisa Kebutuhan Data

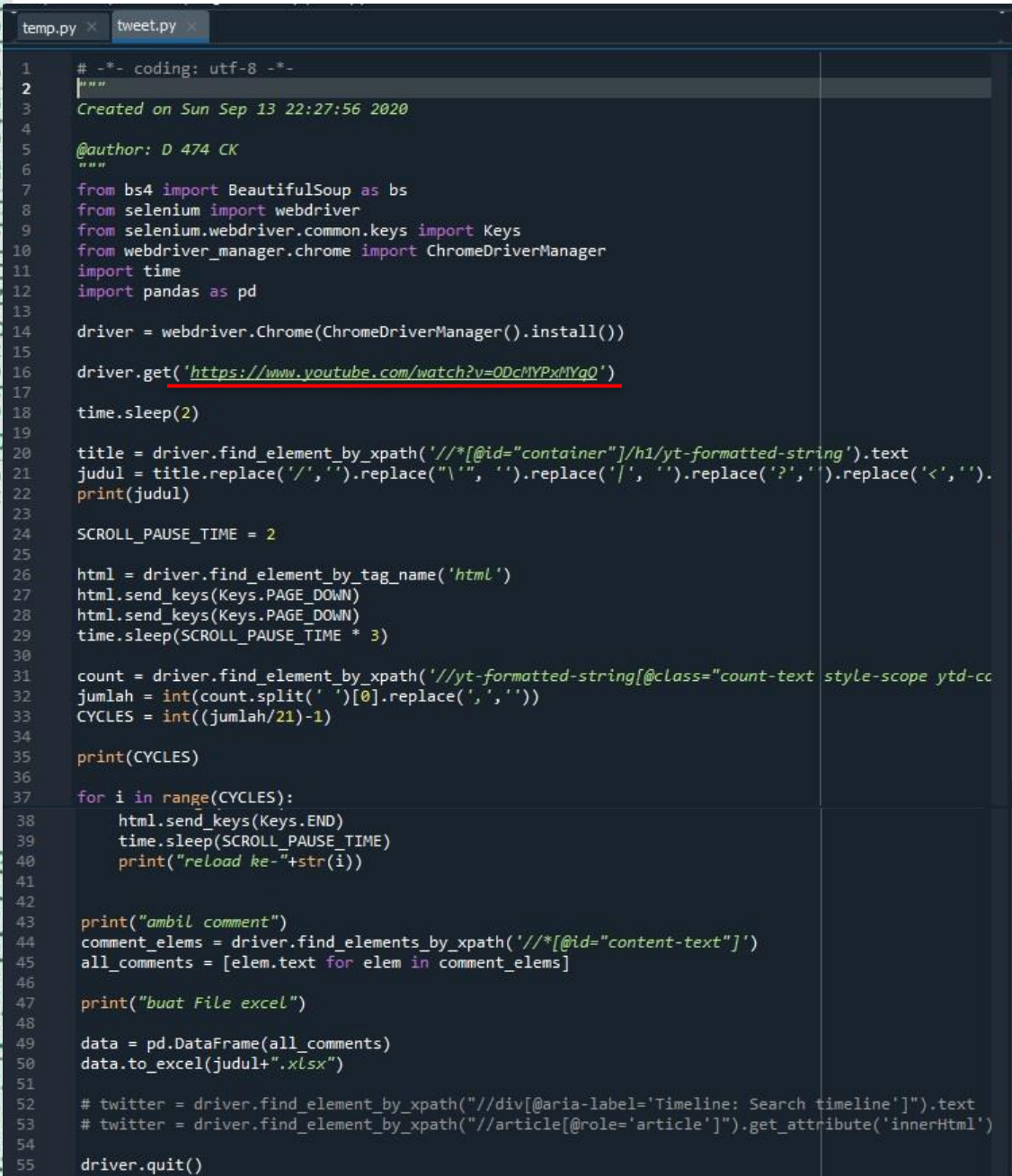
Data yang dibutuhkan pada penelitian ini bersumber dari video *YouTube* berupa komentar pada akun *official channel* Ustad Abdul Somad. Data yang digunakan sebagai dataset yaitu berjumlah 1000 komentar yang akan dibagi menjadi 2 kelas yaitu 500 positif dan 500 negatif. Setelah proses *scraping* data komentar selesai, maka dilanjutkan dengan poses pelabelan manual pada komentar-komentar tersebut.

1. Analisa Pengumpulan Data

Dalam proses pengumpulan data, pengambilan data komentar dilakukan dari tanggal 2 November sampai dengan 16 November 2020 yang diambil dari 3 video pada *channel official* Ustad Abdul Somad dengan menggunakan proses *scraping* pada tools *Spyder*. Dengan cara meletakkan alamat *link* video yang akan di *scraping* untuk proses pengunduhan data komentar hingga proses *scraping* komentar selesai 100% dan data yang digunakan hanya berupa data komentar atau *comment text* saja.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



```

temp.py x tweet.py x
1 # -*- coding: utf-8 -*-
2 """
3 Created on Sun Sep 13 22:27:56 2020
4
5 @author: D 474 CK
6 """
7 from bs4 import BeautifulSoup as bs
8 from selenium import webdriver
9 from selenium.webdriver.common.keys import Keys
10 from webdriver_manager.chrome import ChromeDriverManager
11 import time
12 import pandas as pd
13
14 driver = webdriver.Chrome(ChromeDriverManager().install())
15
16 driver.get('https://www.youtube.com/watch?v=0DcMYPxMYq0')
17
18 time.sleep(2)
19
20 title = driver.find_element_by_xpath('//*[@id="container"]/h1/yt-formatted-string').text
21 judul = title.replace('/', '').replace("\", '').replace('/', '').replace('?', '').replace('<', '').
22 print(judul)
23
24 SCROLL_PAUSE_TIME = 3
25
26 html = driver.find_element_by_tag_name('html')
27 html.send_keys(Keys.PAGE_DOWN)
28 html.send_keys(Keys.PAGE_DOWN)
29 time.sleep(SCROLL_PAUSE_TIME * 3)
30
31 count = driver.find_element_by_xpath('//*[@yt-formatted-string[@class="count-text style-scope ytd-cc
32 jumlah = int(count.split(' ')[0].replace(',',''))
33 CYCLES = int((jumlah/21)-1)
34
35 print(CYCLES)
36
37 for i in range(CYCLES):
38     html.send_keys(Keys.END)
39     time.sleep(SCROLL_PAUSE_TIME)
40     print("reload ke-"+str(i))
41
42
43 print("ambil comment")
44 comment_elems = driver.find_elements_by_xpath('//*[@id="content-text"]')
45 all_comments = [elem.text for elem in comment_elems]
46
47 print("buat File excel")
48
49 data = pd.DataFrame(all_comments)
50 data.to_excel(judul+".xlsx")
51
52 # twitter = driver.find_element_by_xpath("//div[@aria-label='Timeline: Search timeline']").text
53 # twitter = driver.find_element_by_xpath("//article[@role='article']").get_attribute('innerHTML')
54
55 driver.quit()
    
```

Gambar 4.3 Source Code Proses Scraping Data

Berikut daftar alamat *link* video yang digunakan pada penelitian ini.

1. Makna Dibalik Do'a Qunut | Kajian Subuh di Masjid Raya Telun Kenas, Sumut - 9 Agustus 2020.
<https://www.youtube.com/watch?v=A-R338-0088>

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Penonton Film Drama Korea Kafir | Ustad Abdul Somad – UAS
<https://www.youtube.com/watch?v=VHlmyJxcO3M>
3. Ahok Bongkar Kebohongan Ust. Abdul Somad, Kritik Kecurangan UAS – Rizieq Shihab
<https://www.youtube.com/watch?v=ODcMYPxMYqQ>

4.1.2 Pelabelan Manual

Selanjutnya yaitu tahapan Proses pelabelan data komentar pada dataset yang bertujuan untuk menghindari *redundancy* atau data komentar yang bersifat ganda. Proses pelabelan akan dilakukan secara manual oleh ahli bahasa yaitu seorang Dosen Fakultas Ilmu Budaya Universitas Lancang Kuning, Pekanbaru. Proses pelabelan dilakukan untuk membagi data ke dalam 2 kelas yaitu positif dan negatif serta juga membagi dataset menjadi 2 bagian yaitu data latih dan data uji. Data latih adalah data yang digunakan untuk melatih mesin agar dapat mengenali pola, sedangkan data uji adalah data yang digunakan untuk menguji hasil dari pelatihan yang telah dilakukan terhadap mesin. Dibawah ini adalah contoh data komentar yang sudah dikumpulkan dan diberi pelabelan secara manual.

Tabel 4.1 Data Komentar

No	Komentar	Kelas
1.	Alhamdulillah sangat jelas dan manfaat, semoga rahmat dan barokah serta hikmah Allah senantiasa menaungi Ustad Abdul Shomad dalam perjuangannya membawakan firman dan Asma Allah	Positif
2.	Ya Allah setiap harinya saya selalu menonton ceramahnya Ustad Abdul Somad melalui YouTube, dari saya tidak paham sampai saya paham, ilmu yang beliau sampaikan sederhana dan mudah dimengerti bagi kami kami yg sudah lanjut usianya, Amin semoga tuan guru kami selalu di beri kesehatan dan keberkahan	Positif
3.	Ustad Somad juga penistaan agama telah menghina Salib yang menjadi lambang Kemuliaan umat Kristen, hanya kami tidak mau permasalahan, hanya Tuhan Hakim yang Agung yang tau segala sepak terjangmu Somad	Negatif
4.	satu jariunjuk Ahok, tigas jariunjuk kau Somad. senjata makan tuan, selamat nginep di hotel prodeo kau somadddd	Negatif

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.1.3 Pembagian Data

Pada tahapan ini dilakukan pembagian keseluruhan data komentar yang berjumlah 1000 dataset. Data komentar tersebut akan terbagi menjadi 2 bagian yaitu data latih dan data uji yang akan digunakan ke dalam 3 skenario perbandingan data yaitu menjadi 70% data latih dan 30% data uji, 80% data latih dan 20% data uji, 90% data latih dan 10% data uji. Pada pembagian dataset 70:30, data latih akan berjumlah 700 dan data uji berjumlah 300. Pada dataset 80:20, data latih berjumlah 800 dan data uji berjumlah 200. Dan pada pembagian dataset 90:10, data latih berjumlah 900 dan data uji berjumlah 100.

Tabel 4.2 Skenario Pembagian Data

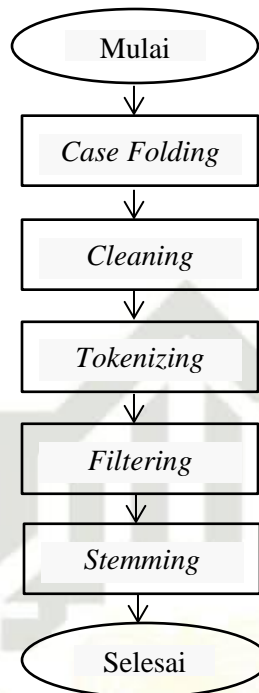
Data Komentar	Kelas	Pembagian Data Latih dan Data uji					
		Latih	Uji	Latih	Uji	Latih	Uji
		70%	30%	80%	20%	90%	10%
1000	Positif	350	150	400	100	450	50
	Negatif	350	150	400	100	450	50
Jumlah		700	300	800	200	900	50

4.1.4 Preprocessing

Setelah melakukan proses pelabelan, data akan melewati tahapan yang namanya *preprocessing*. *Preprocessing* merupakan langkah penting dalam melakukan analisa klasifikasi yang bertujuan untuk membersihkan data dari unsur yang tidak berguna serta tujuan dilakukannya proses ini adalah agar data lebih mudah untuk dipahami oleh sistem pengolahan *text mining*. Adapun tahapan *preprocessing* yang akan dilakukan dalam penelitian ini adalah:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.4 Tahapan *Preprocessing*

Berikut adalah penjelasan dari tahapan *preprocessing* yang disebutkan diatas.

1. Case Folding

Pada tahapan ini teks data komentar akan diubah keseluruhannya menjadi huruf kecil.

Tabel 4.3 Hasil case folding

No	Komentar	Hasil case folding
1.	Alhamdulillah sangat jelas dan manfaat, semoga rahmat dan barokah serta hikmah Allah senantiasa menaungi Ustad Abdul Shomad dalam perjuangannya membawakan firman dan Asma Allah	alhamdulillah sangat jelas dan manfaat, semoga rahmat dan barokah serta hikmah allah senantiasa menaungi ustad abdul shomad dalam perjuangannya membawakan firman dan asma allah
2.	Ya Allah setiap harinya saya selalu menonton ceramahnya Ustad Abdul Somad melalui YouTube, dari saya tidak paham sampai saya paham, ilmu yang beliau sampaikan sederhana dan mudah dimengerti bagi kami kami yang sudah lanjut usianya, Amin semoga tuan guru kami selalu di beri kesehatan dan keberkahan	ya allah setiap harinya saya selalu menonton ceramahnya ustad abdul somad melalui youtube, dari saya tidak paham sampai saya paham, ilmu yang beliau sampaikan sederhana dan mudah dimengerti bagi kami kami yang sudah lanjut usianya, amin semoga tuan guru kami selalu di beri kesehatan dan keberkahan
3.	Ustad Somad juga penistaan agama telah	ustad somad juga penistaan agama telah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	menghina Salib yang menjadi lambang Kemuliaan umat Kristen , hanya kami tidak mau permasalahan, hanya Tuhan Hakim yang Agung yang tau segala sepak terjangmu Somad	menghina salib yang menjadi lambang kemuliaan umat kristen, hanya kami tidak mau permasalahan, hanya tuhan hakim yang agung yang tau segala sepak terjangmu somad
4.	satu jari nunjuk Ahok , tiga jari nunjuk kau Somad . Senjata makan tuan, selamat ngingep di hotel prodeo somadddd	satu jari nunjuk ahok, tiga jari nunjuk kau somad. senjata makan tuan, selamat ngingep di hotel prodeo somadddd

2. Cleaning

Tahapan *cleaning* dilakukan untuk mengurangi kata yang dianggap tidak diperlukan pada teks dengan cara menghapus atau menghilangkan simbol, angka, tanda baca, URL, *emoticon* dan *mention* (@).

Tabel 4.4 Hasil *cleaning*

No	Komentar	Hasil <i>cleaning</i>
1.	alhamdulillah sangat jelas dan manfaat, semoga rahmat dan barokah serta hikmah allah senantiasa menaungi ustad abdul shomad dlm perjuangannya membawakan firman dan asma allah	alhamdulillah sangat jelas dan manfaat semoga rahmat dan barokah serta hikmah allah senantiasa menaungi ustad abdul shomad dlm perjuangannya membawakan firman dan asma allah
2.	ya allah setiap harinya saya selalu menonton ceramahnya ustad abdul somad melalui youtube, dari saya tidak paham sampai saya paham, ilmu yang beliau sampaikan sederhana dan mudah dimengerti bagi kami kami yang sudah lanjut usianya, amin semoga tuan guru kami selalu di beri kesehatan dan keberkahan	ya allah setiap harinya saya selalu menonton ceramahnya ustad abdul somad melalui youtube dari saya tidak paham sampai saya paham ilmu yang beliau sampaikan sederhana dan mudah dimengerti bagi kami kami yang sudah lanjut usianya amin semoga tuan guru kami selalu di beri kesehatan dan keberkahan
3.	ustad somad juga penistaan agama telah menghina salib yang menjadi lambang kemuliaan umat kristen, hanya kami tidak mau permasalahan, hanya tuhan hakim yang agung yang tau segala sepak terjangmu somad	ustad somad juga penistaan agama telah menghina salib yang menjadi lambang kemuliaan umat kristen hanya kami tidak mau permasalahan hanya tuhan hakim yang agung yang tau segala sepak terjangmu somad
4.	sat jari nunjuk ahok, tiga jari nunjuk kau somad, senjata makan tuan, selamat ngingep di hotel prodeo somadddd	satu jari nunjuk ahok tiga jari nunjuk kau somad senjata makan tuan selamat ngingep di hotel prodeo somadddd

3. Tokenizing

Tahapan *Tokenizing* adalah tahapan pemisahan teks menjadi potongan-potongan kata yang disebut sebagai token untuk kemudian di analisa.

Tabel 4.5 Hasil *tokenizing*

D1	D2	D3	D4
alhamdulillah	ya	ustad	satu
sangat	allah	somad	jari
jelas	setiap	juga	tunjuk
dan	harinya	penistaan	ahok
manfaat	saya	agama	tiga
semoga	selalu	telah	jari
rahmat	menonton	menghina	tunjuk
dan	ceramahnya	salib	kau
barokah	ustad	yang	somad
serta	abdul	menjadi	senjata
hikmah	somad	lambang	makan
allah	melalui	kemuliaan	tuan
senantiasa	youtube	umat	selamat
menaungi	dari	kristen	nginap
ustad	saya	hanya	di
abdul	tidak	kami	hotel
somad	paham	tidak	prodeo
dalam	sampai	mau	somaddddd
perjuangan	saya	permasalahkan	
nya	paham	hanya	
membawakan	ilmu	tuhan	
firman	yang	hakim	
dan	beliau	yang	
asma	sampaikan	agung	
allah	sederhana	yang	
	dan	tahu	
	mudah	segala	
	dimengerti	sepak	
	bagi	terjangmu	
	kami	somad	
	kami		
	yang		
	sudah		
	lanjut		
	usianya		
	amin		
	semoga		
	tuan		
	guru		
	kami		
	selalu		
	di		

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikis kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	beri		
	kesehatan		
	dan		
	keberkahan		

4. Filtering

Tahapan *Filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing* sebelumnya. Bisa menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan sebagainya

Tabel 4.6 Hasil filtering

D1	D2	D3	D4
alhamdullillah	allah	ustad	satu
Jelas	harinya	somad	jari
manfaat	saya	juga	tunjuk
rahmat	menonton	penistaan	ahok
berkah	ceramahnya	agama	tiga
hikmah	ustad	menghina	jari
allah	abdul	salib	tunjuk
senantiasa	somad	lambang	kau
menaungi	youtube	kemuliaan	somad
ustad	saya	umat	senjata
abdul	tidak	kristen	makan
somad	paham	kami	tuan
perjuangan	sampai	permasalahkan	selamat
membawakan	saya	tuhan	nginap
firman	paham	hakim	hotel
asma	ilmu	agung	prodeo
allah	beliau	tahu	somad
	sampaikan	segala	
	sederhana	sepak	
	mudah	terjangmu	
	dimengerti	somad	
	kami		
	kami		
	sudah		
	lanjut		
	usianya		
	amin		
	semoga		
	tuan		

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	guru		
	kami		
	selalu		
	beri		
	kesehatan		
	keberkahan		

5. Stemming

Pada tahapan *Stemming* ini yaitu mencari kata dasar dari setiap kata hasil filtering sebelumnya. *Stemming* akan melakukan proses pengembalian berbagai bentukan kata menjadi bentukan kata dasar dengan menghilangkan imbuhan baik awalan dan akhiran. Proses *stemming* dilakukan dengan cara meng-import kelas *StemmerFactory* dari *library* sastrawi dengan menggunakan algoritma Nazief & Adriani. Berikut merupakan hasil data komentar yang telah di *stemming*.

Tabel 4.7 Hasil stemming

d1	d2	d3	d4
alhamdulillah	allah	ustad	satu
jelas	hari	somad	jari
manfaat	saya	juga	tunjuk
rahmat	tonton	nista	ahok
berkah	ceramah	agama	tiga
hikmah	ustad	hina	jari
allah	abdul	salib	tunjuk
senantiasa	somad	lambang	kau
naung	youtube	mulia	somad
ustad	saya	umat	senjata
abdul	tidak	kristen	makan
somad	paham	kami	tuan
juang	sampai	masalah	selamat
bawa	saya	tuhan	inap
firman	paham	hakim	hotel
asma	ilmu	agung	prodeo
allah	beliau	tahu	somad
	sampai	segala	
	sederhana	sepak	
	mudah	terjang	
	mengerti	somad	
	kami		
	kami		
	sudah		
	lanjut		

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	usia		
	amin		
	semoga		
	tuan		
	guru		
	kami		
	selalu		
	beri		
	sehat		
	berkah		

4.1.5 Term Frequency – Inverse Document Frequency (TF-IDF)

Setelah melewati tahapan *preprocessing*, tahapan selanjutnya yang akan dilakukan adalah pemberian bobot pada setiap kata yang ada pada data komentar dengan menggunakan TF-IDF. *Term Frequency* (TF) yang akan digunakan untuk menghitung seberapa sering sebuah kata muncul pada data komentar dan *inverse document frequency* (IDF) untuk pemberian bobot pada kata tertentu yang banyak terkandung di dalam sebuah dokumen. (Persamaan 2.2)

Tabel 4.8 Hasil TF-IDF

Term (t)	TF				DF	IDF	TF-IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
alhamdulillah	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
jelas	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
manfaat	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
rahmat	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
berkah	1	1	0	0	2	Log(4/2)=0.301	0.301	0.301	0	0
hikmah	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
allah	2	1	0	0	2	Log(4/2)=0.301	0.602	0.301	0	0
senantiasa	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
naung	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
ustad	1	1	1	0	3	Log(4/3)=0.124	0.124	0.124	0.124	0
abdul	1	1	0	0	2	Log(4/2)=0.301	0.301	0.301	0	0
somad	1	1	2	2	4	Log(4/4)=0	0	0	0	0
juang	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
bawa	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
firman	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
asma	1	0	0	0	1	Log(4/1)=0.602	0.602	0	0	0
hari	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
saya	0	3	0	0	1	Log(4/1)=0.602	0	1.806	0	0
tonton	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
ceramah	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

youtube	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
tidak	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
paham	0	2	0	0	1	Log(4/1)=0.602	0	1.204	0	0
sampai	0	2	0	0	1	Log(4/1)=0.602	0	1.204	0	0
ilmu	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
beliau	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
sederhana	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
mudah	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
mengerti	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
kami	0	3	1	0	2	Log(4/2)=0.301	0	0.903	0.301	0
sudah	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
lanjut	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
usia	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
amin	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
semoga	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
tuan	0	1	0	1	2	Log(4/2)=0.301	0	0.301	0	0.301
guru	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
selalu	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
beri	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
sehat	0	1	0	0	1	Log(4/1)=0.602	0	0.602	0	0
juga	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
nista	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
agama	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
hina	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
salib	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
lambang	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
mulia	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
umat	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
kristen	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
masalah	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
tuhan	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
hakim	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
agung	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
tahu	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
segala	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
sepak	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
terjang	0	0	1	0	1	Log(4/1)=0.602	0	0	0.602	0
satu	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602
jari	0	0	0	2	1	Log(4/1)=0.602	0	0	0	1.204
tunjuk	0	0	0	2	1	Log(4/1)=0.602	0	0	0	1.204
ahok	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602
tiga	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602
kau	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602
senjata	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602
makan	0	0	0	1	1	Log(4/1)=0.602	0	0	0	0.602

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

selamat	0	0	0	1	1	$\text{Log}(4/1)=0.602$	0	0	0	0.602
inap	0	0	0	1	1	$\text{Log}(4/1)=0.602$	0	0	0	0.602
hotel	0	0	0	1	1	$\text{Log}(4/1)=0.602$	0	0	0	0.602
prodeo	0	0	0	1	1	$\text{Log}(4/1)=0.602$	0	0	0	0.602
Total							7.950	15.174	10.659	8.729
							42.512			

4.1.6 Tahap Klasifikasi *Support Vector Machine* (SVM)

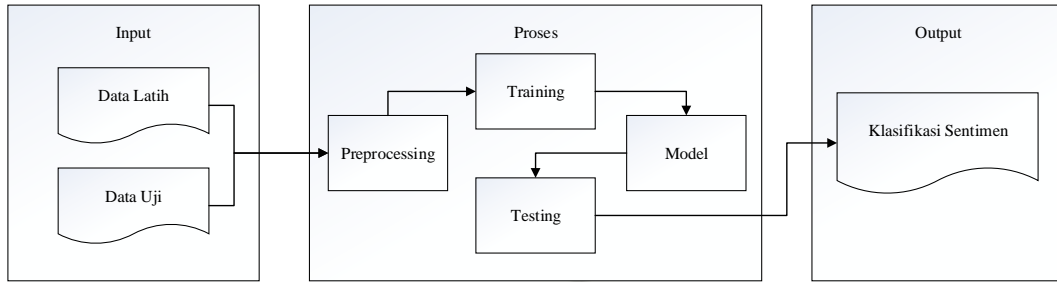
Proses SVM akan melibatkan *training* menggunakan data latih untuk menghasilkan model pembelajaran dengan menggunakan *k-fold cross validation*, serta menggunakan parameter C dan *Gamma*. Model terbaik merupakan model yang memiliki akurasi nilai tertinggi dari pasangan C dan *Gamma* terbaik yang diperoleh dari Teknik *grid searching* dan dengan menggunakan *10-fold cross validation*. Pada proses *testing* menggunakan data uji untuk mengklasifikasikan data uji tersebut.

1. Pembelajaran Dan Model

Pembelajaran pada SVM menggunakan data latih untuk mendapatkan model klasifikasi pada SVM. Tahap pembelajaran ini melibatkan fungsi kernel sebagai fungsi transformasi. Kernel yang digunakan adalah kernel RBF. Kernel RBF membutuhkan pasangan parameter C dan *Gamma*. Untuk mendapatkan nilai parameter C dan *Gamma* terbaik dilakukan dengan metode *grid search* dan *10-fold cross validation*. Metode *grid search* bertujuan untuk membuat *grid* parameter dengan cara menentukan nilai untuk parameter C dan *Gamma* secara manual. Menentukan nilai parameter C dan menentukan nilai parameter *Gamma*, dan melakukan pencarian *grid* untuk C dan *Gamma* dengan cara menerapkan *10-fold cross validation*. *10-fold cross validation* adalah pembagian data latih menjadi 10 segmen sama banyak. Kemudian akan dilakukan 10 kali proses *training* dan *testing* dengan perbandingan 9/10 segmen sebagai data latih dan 1/10 segmen sebagai data uji. Selanjutnya menghitung rata-rata akurasi untuk keseluruhan *fold*. Kemudian melakukan pemilihan parameter C dan *Gamma* terbaik berdasarkan akurasi paling tinggi. Berikut merupakan tahapan klasifikasi SVM secara umum dapat dilihat pada gambar dibawah.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.5 Tahapan Klasifikasi SVM

Pada gambar 4.4 ada tiga proses yang dilakukan, yaitu:

1. *Input*

Data yang akan diinputkan adalah keseluruhan data komentar yang telah dikumpulkan sebanyak 1000 dan telah dilakukan proses pelabelan secara manual dengan bantuan ahli bahasa. Data akan terbagi menjadi beberapa skenario yang akan menghasilkan validasi akurasi tertinggi

2. Proses SVM

Merupakan tahapan mulai dari *preprocessing*, pembobotan sampai pembentukan pemodelan. Dalam proses ini juga sudah dilakukan validasi akurasi menggunakan *k-fold* dan *confusion matrix*.

3. *Output*

Data yang akan dihasilkan adalah berupa data yang telah diprediksi oleh pemodelan svm dengan kelas positif dan negatif.

4.2 Perancangan

Tahapan perancangan menampilkan *pseudocode* program klasifikasi sentimen menggunakan *support vector machine* serta perancangan eksperimen yang dilakukan pada penelitian ini.

4.2.1 Pseudocode Scraping Data

Program Scraping Data

```

from bs4 import BeautifulSoup as bs
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from webdriver_manager.chrome import ChromeDriverManager
import time
  
```

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

import pandas as pd

driver ← webdriver.Chrome(ChromeDriverManager().install())

driver.get('https://www.youtube.com/watch?v=ODcMYPxMYqQ')

time.sleep(2)

title ← driver.find_element_by_xpath('//*[@id="container"]/h1/yt-formatted-string').text
judul ← title.replace('/', '').replace("\", '').replace('|', '').replace('?', '').Replace('<', '').Replace('>', '').replace("'", '')print(judul)

SCROLL_PAUSE_TIME ← 2

html ← driver.find_element_by_tag_name('html')
html.send_keys(Keys.PAGE_DOWN)
html.send_keys(Keys.PAGE_DOWN)
time.sleep(SCROLL_PAUSE_TIME * 3)

count ← driver.find_element_by_xpath('//*[@yt-formatted-string[@class="count-text style-scope ytd-comments-header-renderer"]').text
jumlah ← int(count.split(' ')[0].replace(',',''))
CYCLES ← int((jumlah/21)-1)

print(CYCLES)

for i in range(CYCLES):
html.send_keys(Keys.END)
time.sleep(SCROLL_PAUSE_TIME)
print("reload ke-"+str(i))

print("ambil comment")
comment_elems ← driver.find_elements_by_xpath('//*[@id="content-text"'])
all_comments ← [elem.text for elem in comment_elems]

print("buat File excel")

data ← pd.DataFrame(all_comments)
data.to_excel(judul+".xlsx")

driver.quit()

```



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.2.2 Pseudocode Preprocessing Data

Program *preprocessing_data*

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory ← StemmerFactory()
stemmer ← factory.create_stemmer()
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory
factory ← StopWordRemoverFactory()
stopwords ← factory.get_stop_words()
# print(stopwords)
stopword ← factory.create_stop_word_remover()
kalimat ← df["txt"]
kalimat ← kalimat.transpose()
hasil ← []
casefolding ← []
cleaning ← []
tokenizing ← []
steming ← []

preprocessing ← []
# print(df)
for i in range(len(df)):
    preprocessing[i:i] ← [[kalimat[i],
                           kalimat[i].lower(),
                           kalimat[i].lower().translate(str.maketrans("", "", string.punctuation)),
                           stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation))),
                           stopword.remove(stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation)))),
                           stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation))).split()
                           ]]

    kalimat[i] ← kalimat[i].translate(str.maketrans("", "", string.punctuation))
    kalimat[i] ← stemmer.stem(kalimat[i])
##df[i] ← df[i].split()
    hasil[i:i] ← [stopword.remove(kalimat[i])]
#hasil[i:i] ← nltk.tokenize.word_tokenize(df[i])

preprocessing ← np.array(preprocessing)
preprocessing ← pd.DataFrame({'Data Komentar': preprocessing[:,0]
},

```


Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikis kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```
'Sesudah Case Folding':preprocessing[:,
1],
'Sesudah Cleaning':preprocessing[:, 2],
'Sesudah Stemming':preprocessing[:, 3],
'Sesudah Stopword':preprocessing[:, 4],
'Sesudah Tokenizing':preprocessing[:, 5]
})
```

```
print(preprocessing)
```

4.2.3 Pseudocode TF-IDF

```
#in this case our d
y <- df["liked"]
#convert df.txt from text to features
vectorizer <- TfidfVectorizer()
X <- vectorizer.fit_transform(hasil)
# print(vectorizer.get_feature_names())
# print(X.shape)
# X <- vectorizer.fit_transform(df.txt)
#6918 observations x 2022 unique words.
print (y.shape)
print (X.shape)
```

```
#Test Train Split as usual
X_train, X_test,y_train, y_test <- train_test_split(X, y, r
andom_state <- 2)
```

4.2.4 Pseudocode Support Vector Machine

Input

```
# defining parameter range
param_grid = {'C': [0.1, 1, 10, 100, 1000],
'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
'kernel': ['rbf']}
```

```
grid = GridSearchCV(SVC(), param_grid, refit = True, verbo
se = 3)
```

```
# fitting the model for grid search
grid.fit(X_train, y_train)
```

```
# print best parameter after tuning
print(grid.best_params_)
```

```
# print how our model looks after hyper-parameter tuning
print(grid.best_estimator_)
```

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Output

```
{'C': 10, 'gamma': 1, 'kernel': 'rbf'}
SVC(C=10, break_ties=False, cache_size=200,
    class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=1,
    kernel='rbf', max_iter=-1, probability=False,
    random_state=None, shrinking=True, tol=0.001,
    verbose=False)
```



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB V

IMPLEMENTASI DAN PENGUJIAN

5.1 Implementasi

Implementasi merupakan tahap penerapan analisa pemodelan yang sudah dilakukan berdasarkan hasil perancangan yang telah di *design* sebelumnya. Implementasi dilakukan menggunakan *source code* bahasa pemrograman *python* untuk menjalankan model.

5.1.1 Batasan Implementasi

Implementasi penerapan metode *support vector machine* dalam melakukan klasifikasi sentimen terhadap Ustad Abdul Somad memiliki batasan sebagai berikut:

1. Menggunakan bahasa pemrograman *Python*.
2. Dataset yang digunakan berjumlah 1000 data komentar.
3. Dataset akan terbagi kedalam 2 kelas, yaitu positif dan negatif.

5.1.2 Lingkungan Implementasi

Membangun dan menjalankan sebuah model *text mining* akan dapat dilakukan dengan ketersediaan perangkat keras (*hardware*) yang handal. Dan selain itu juga diperlukan bantuan dari perangkat lunak (*software*) sebagai sarana dalam melakukan proses pemodelan.

1. *Hardware* yang digunakan dengan spesifikasi:

Processor : Intel(R) Core(TM) i3 CPU M 380 @2.53GHz
Memory (RAM) : 4.00 GB (3.80 GB usable)
HDD : 500 GB

2. *Software* yang digunakan adalah:

Sistem Operasi : *Windows 7 Ultimate*
Code Implementation : *Spyder (Anaconda3)* dan *Google Colab*
Browser : *Google Chrome 87.0.4*
Bahasa Pemrograman : *Python 3.7.6*
Library : *nlk, matplotlib, numpy, pandas*, dan *sastrawi*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.1.3 Implementasi Pemodelan

Tahapan implementasi pemodelan dilakukan menggunakan perintah kode *python*. Proses pengkodean dilakukan bertahap menggunakan *google colab*. Berikut ini tahapan-tahapan kodingan yang dilakukan dalam pengimplementasian pemodelan.

1. Implementasi *Load Library Python*

Tahapan pertama dalam proses pengkodean pemodelan adalah *load library python*. Untuk menggunakan *library python* digunakan perintah *import*. Berikut *library python* yang digunakan pada penelitian ini.

```
import pandas as pd
import numpy as np
import nltk
import csv
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
from sklearn.metrics import confusion_matrix
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.svm import SVC
from sklearn.metrics import roc_auc_score
from sklearn import datasets
from sklearn.preprocessing import StandardScaler
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import string
from nltk.tokenize import word_tokenize
```

Gambar 5.1 *Load Library Python*

2. Implementasi *Input Data Komentar*

Tahapan memasukkan data komentar kedalam model melalui *file directory*.

```
!ls '/content/drive/My Drive/New'
file= '/content/drive/My Drive/New/datauji100.csv'
file1= '/content/drive/My Drive/New/utube.csv'
```

Gambar 5.2 *Load File Data Komentar*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Implementasi *Dataframe* dan Pelabelan Data

Data komentar yang telah berhasil di scrapping kemudian dilabelkan masih dalam berbentuk file *excel*. Oleh karena itu perlu untuk dikonversikan ke dalam *dataframe*.



	label	komentar
772	negatif	Ustad, kok begitu ya
437	positif	Semoga USTADZ selalu sehat Dan diRidhoi ALLOH ...
847	negatif	Itu ustadz pemuka agama, hati" adzab kubur bang
880	negatif	aduh ustad somad kok ngomong sembarangan ya mi...
116	positif	Alhamdulillah. saya senang banget dengar cera...
...
125	positif	Semoga ustad sehat sejahtera selalu. Aamiin
31	positif	Ustad Abdul Somad mukanya bersinar, cerah, m...
873	negatif	Ustad Abdul Somad ustad yang cuma pinter Teori...
335	positif	2 orang berilmu tinggi di indonesia. Di sumat...
955	negatif	Somat gembeng

1000 rows × 2 columns

Gambar 5.3 *Dataframe* Pelabelan Data

4. Implementasi *Text Preprocessing* Data Komentar

Pada tahapan ini akan dijelaskan implementasi hasil perancangan dari proses *preprocessing text* yang meliputi *case folding*, *cleaning*, *stemming*, *filtering stopword*, dan *tokenizing* yang dilakukan menggunakan *tools Jupiter Notebook*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```
#TEXT PREPROCESSING

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
# print(stopwords)
stopword = factory.create_stop_word_remover()
kalimat = df["txt"]
kalimat = kalimat.transpose()
hasil = []
casefolding=[]
cleaning=[]
tokenizing = []
steming = []

preprocessing = []
# print(df)
for i in range(len(df)):
    preprocessing[i:i] = [[kalimat[i],
                          kalimat[i].lower(),
                          kalimat[i].lower().translate(str.maketrans("", "", string.punctuation)),
                          stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation))),
                          stopword.remove(stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation))),
                          stemmer.stem(kalimat[i].lower().translate(str.maketrans("", "", string.punctuation))).split()
                          ]]
    kalimat[i] = kalimat[i].translate(str.maketrans("", "", string.punctuation))
    kalimat[i] = stemmer.stem(kalimat[i])
# # df[i] = df[i].split()
hasil[i:i] = [stopword.remove(kalimat[i])]
# hasil[i:i] = nltk.tokenize.word_tokenize(df[i])

preprocessing = np.array(preprocessing)
preprocessing = pd.DataFrame({'Data Komentar': preprocessing[:, 0],
                             'Sesudah Case Folding': preprocessing[:, 1],
                             'Sesudah Cleaning': preprocessing[:, 2],
                             'Sesudah Stemming': preprocessing[:, 3],
                             'Sesudah Stopword': preprocessing[:, 4],
                             'Sesudah Tokenizing': preprocessing[:, 5]
                             })

print(preprocessing)

preprocessing.to_csv("tube.csv")
```

Gambar 5.4 Source Code Text Preprocessing

Pada tampilan hasil *preprocessing* juga terdapat tampilan keseluruhan *preprocessing* data komentar yang berjumlah sebanyak 1000 komentar. Yang ditampilkan hanya 5 bagian diatas dan 5 bagian dibawah dataset. Dan juga terdapat tampilan dari *case folding*, yaitu merubah teks data komentar kedalam bentuk huruf kecil semua.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

Data Komentar \
0 Alhamdulillah dengar ceramah Ustad Abdul Somad ...
1 Kapan aku bisa ketemu sama Ustad Abdul Somad ...
2 mohon dukungannya saudara saudariku untuk bant...
3 semoga Almkarram Tuan guru ustad Abdul Shomad...
4 Semoga Allah menyembatkan & memanjang usia beli...
..
...
995 Ustad tomad ini lucu semoga gilamu cepat sembu...
996 Wah ternyata iblis nya Abdul Somad sendiri yan...
997 Ustat somat itu benar dan jelas menghina orang...
998 jangan penah menjelek jelekkan agama orang lai...
999 Sebaiknya Ustad Abdul Somad aplikasikan aja ....

Sesudah Case Folding \
0 alhamdulillah dengar ceramah ustad abdul somad ...
1 kapan aku bisa ketemu sama ustad abdul somad ...
2 mohon dukungannya saudara saudariku untuk bant...
3 semoga almkarram tuan guru ustad abdul shomad...
4 semoga allah menyembatkan & memanjang usia beli...
..
...
995 ustad tomad ini lucu semoga gilamu cepat sembu...
996 wah ternyata iblis nya abdul somad sendiri yan...
997 ustat somat itu benar dan jelas menghina orang...
998 jangan penah menjelek jelekkan agama orang lai...
999 sebaiknya ustad abdul somad aplikasikan aja ....

```

Gambar 5.5 Tampilan Data Komentar dan Hasil *Case Folding*

Selanjutnya adalah tampilan *cleaning* dan *stemming*. Yaitu *cleaning* adalah tahapan menghapus berbagai tanda baca di dalam suatu teks. Dan *stemming* adalah proses tahapan merubah suatu kata ke dalam bentuk kata dasarnya.

```

Sesudah Cleaning \
0 alhamdulillah dengar ceramah ustad abdul somad ...
1 kapan aku bisa ketemu sama ustad abdul somad ...
2 mohon dukungannya saudara saudariku untuk bant...
3 semoga almkarram tuan guru ustad abdul shomad...
4 semoga allah menyembatkan memanjang usia belia...
..
...
995 ustad tomad ini lucu semoga gilamu cepat sembu...
996 wah ternyata iblis nya abdul somad sendiri yan...
997 ustat somat itu benar dan jelas menghina orang...
998 jangan penah menjelek jelekkan agama orang lai...
999 sebaiknya ustad abdul somad aplikasikan aja c...

Sesudah Stemming \
0 alhamdulillah dengar ceramah ustad abdul somad ...
1 kapan aku bisa ketemu sama ustad abdul somad y...
2 mohon dukung saudara saudari untuk bantu subsc...
3 moga almkarram tuan guru ustad abdul shomad d...
4 moga allah sehat panjang usia beliau kami masi...
..
...
995 ustad tomad ini lucu moga gila cepat sembuh da...
996 wah nyata iblis nya abdul somad sendiri yang p...
997 ustat somat itu benar dan jelas hina orang kri...
998 jangan penah jelek jelek agama orang lain usta...
999 baik ustad abdul somad aplikasi aja cara milih...

```

Gambar 5.6 Tampilan *Cleaning* dan *Stemming*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Selanjutnya adalah tampilan tahapan *Stopword* dan *tokenizing*. *Stopword* yaitu tahapan menghapus kata-kata yang tidak terlalu memiliki arti terhadap dataset. Dan *tokenizing* adalah tahapan pemotongan data komentar menjadi kumpulan kata.

```

                                Sesudah Stopword \
0  alhamdulillah dengar ceramah ustad abdul somad ...
1  kapan aku ketemu sama ustad abdul somad allah ...
2  mohon dukung saudara saudari bantu subscribe c...
3  moga almukarram tuan guru ustad abdul shomad d...
4  moga allah sehat panjang usia beliau masih san...
..  ...
995 ustad tomad lucu moga gila cepat sembuh bisa m...
996 wah nyata iblis nya abdul somad sendiri penuh ...
997 ustat somat benar jelas hina orang kristen bia...
998 jangan penah jelek jelek agama orang ustad bel...
999 baik ustad abdul somad aplikasi aja cara milih...

                                Sesudah Tokenizing
0  [alhamdulillah, dengar, ceramah, ustad, abdul, ...
1  [kapan, aku, bisa, ketemu, sama, ustad, abdul,...
2  [mohon, dukung, saudara, saudari, untuk, bantu...
3  [moga, almukarram, tuan, guru, ustad, abdul, s...
4  [moga, allah, sehat, panjang, usia, beliau, ka...
..  ...
995 [ustad, tomad, ini, lucu, moga, gila, cepat, s...
996 [wah, nyata, iblis, nya, abdul, somad, sendiri...
997 [ustat, somat, itu, benar, dan, jelas, hina, o...
998 [jangan, penah, jelek, jelek, agama, orang, la...
999 [baik, ustad, abdul, somad, aplikasi, aja, car...
    
```

Gambar 5.7 Tampilan *Stopword* dan *Tokenizing*

5. Implementasi Pembobotan TF-IDF

Tahap pemberian bobot pada setiap kata dengan menghitung *Term Frequency* dan *Inverse Document Frequency*, bisa dilihat pada gambar dibawah.

```

[12] #in this case our d
      y = df["liked"]

[13] #convert df.txt from text to features
      vectorizer = TfidfVectorizer()
      X = vectorizer.fit_transform(hasil)
      # print(vectorizer.get_feature_names())
      # print(X.shape)
      # X= vectorizer.fit_transform(df.txt)

[14] #6918 observations x 2022 unique words.
      print (y.shape)
      print (X.shape)

      (1000,)
      (1000, 3316)

[15] #Test Train Split as usual
      X_train, X_test,y_train, y_test = train_test_split(X, y, random_state=2)
    
```

Gambar 5.8 Implementasi TF-IDF menggunakan Dataset

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengikis kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6. Implementasi Metode *Support Vector Machine*

Implementasi klasifikasi akan menghasilkan suatu hasil klasifikasi data baru yang telah melalui semua proses *preprocessing*. Data tersebut tidak perlu melakukan proses pelabelan manual karena akan di proses oleh SVM *testing* untuk mengelompokkan sesuai dengan hasil klasifikasi pada SVM itu sendiri. Berikut adalah hasil klasifikasi yang berhasil diprediksi oleh sistem.

```

▶ # defining parameter range
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'kernel': ['rbf']}

grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3)

# fitting the model for grid search
grid.fit(X_train, y_train)

▶ # print best parameter after tuning
print(grid.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid.best_estimator_)

{'C': 10, 'gamma': 1, 'kernel': 'rbf'}
SVC(C=10, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=1, kernel='rbf', max_iter=-1,
    probability=False, random_state=None, shrinking=True, tol=0.001,
    verbose=False)

```

Gambar 5.9 Implementasi Pencarian Parameter Terbaik C dan *Gamma*

5.2 Pengujian

Pada tahapan ini dilakukan pengujian pada pemodelan untuk mengukur tingkat akurasi klasifikasi sentimen yang telah dilakukan. Pengujian akurasi dapat dilakukan dengan cara membandingkan persentase hasil klasifikasi yang menggunakan model pembelajaran *support vector machine* dengan label masing-masing kelas yang sudah dilakukan pada setiap data uji.

5.2.1 Pengujian Nilai Parameter C dan *Gamma* Terbaik

Pengujian parameter C dan *gamma* dilakukan untuk mendapatkan nilai C dan *gamma* yang menghasilkan akurasi terbaik. Pengujian dilakukan dengan *range* nilai C =

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

0.1, 1, 10, 100, 1000 dan $\gamma = 1, 0.1, 0.01, 0.001, 0.0001$. Setiap parameter akan di gunakan sesuai dengan parameter kernel yang digunakan yaitu kernel RBF. Dataset yang digunakan berjumlah 1000 komentar. Pengujian pertama diambil 70% sebagai data latih dan 30% sebagai data uji. Pengujian kedua diambil 80% sebagai data latih dan 20% sebagai data uji. Sedangkan untuk pengujian ketiga diambil 90% sebagai data latih dan 10% sebagai data uji.

Tabel 5.1 Pencarian Parameter Terbaik

C/γ	1	0.1	0.01	0.001	0.0001
0.1	51%	51%	51%	51%	51%
1	75%	75%	51%	51%	51%
10	88%	75%	75%	51%	51%
100	75%	75%	75%	75%	51%
1000	75%	75%	75%	75%	75%

Pada tabel diatas dapat dilihat bahwa nilai parameter kernel RBF terbaik terdapat pada nilai $C = 10$ dan nilai $\gamma = 1$ berdasarkan pemodelan yang dilakukan dengan menghasilkan akurasi sebesar 88%.

5.2.2 Pengujian Model Terbaik

Pemilihan model terbaik akan menghasilkan nilai akurasi terbaik yang didapatkan dari model terbaik. Tabel rincian hasil skema *k-fold* yang di hasilkan oleh parameter kernel terbaik adalah sebagai berikut:

Tabel 5.2 Pengujian Model Terbaik Kernel RBF

Fold	Akurasi	Akurasi Testing
-1	100%	97.14%
-2	100%	97.14%
-3	100%	94.28%
-4	100%	94.28%
-5	100%	94.28%
-6	100%	95.71%
-7	100%	95.71%
-8	100%	95.71%
-9	100%	95.71%
-10	100%	91.42%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pada pengujian tabel 5.2 dapat dilihat model terbaik *k-fold cross validation* dengan menggunakan parameter terbaik menghasilkan nilai akurasi testing tertinggi sebesar 97.14% yaitu pada *fold* ke-1.

5.2.3 Pengujian Validasi Akurasi

Dataset yang digunakan adalah sebanyak 1.000 komentar (500 positif dan 500 negatif). Pengujian pertama diambil 10% sebagai data uji. Pengujian kedua diambil 20% sebagai data uji. Untuk pengujian ketiga diambil 30% sebagai data uji. Kernel yang digunakan pada penelitian ini adalah kernel RBF dengan nilai parameter ($C = 10$ dan $\text{Gamma} = 1$)

Tabel 5.3 Perbandingan Akurasi Data Uji

No.	Perbandingan	Data Latih	Data Uji	Akurasi
1.	90:10%	900	100	93.0 %
2.	80:20%	800	200	95.5 %
3.	70:30%	700	300	97.33%

5.2.4 Tujuan Pengujian

Tujuan pengujian dalam penelitian tugas akhir ini adalah untuk mengukur akurasi dari klasifikasi sentimen dari pemodelan yang telah dilakukan pada proses pembelajaran. Pengujian ini menggunakan model *support vector machine* dengan label klasifikasi sentimen yang sebelumnya sudah ada pada data uji. Tujuan selanjutnya adalah agar mampu mengukur tingkat kesalahan dari klasifikasi yang akan dilakukan.

5.3 Evaluasi

Hasil pengujian pertama menggunakan metode *support vector machine* pada data uji sebanyak 10% dengan jumlah data sebanyak 100 komentar, hasil akurasi sebesar 94.0%. Hasil prediksi menggunakan *confusion matrix* dapat dilihat pada tabel berikut.

Tabel 5.4 Hasil *Confusion Matrix* Data Uji 10 %

	<i>True Positive</i>	<i>False Negative</i>
<i>True Positive</i>	43	7
<i>False Negative</i>	0	50

$$\begin{aligned}
 \text{Akurasi} &= \frac{43 + 50}{43 + 7 + 0 + 50} \times 100 \\
 &= \frac{93}{100} \times 100 \\
 &= 93.0 \%
 \end{aligned}$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak mengizinkan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hasil pengujian kedua menggunakan metode *support vector machine* pada data uji sebanyak 20% dengan jumlah data sebanyak 200 komentar, hasil akurasi sebesar 96.5 %. Hasil prediksi menggunakan *confusion matrix* dapat dilihat pada tabel berikut.

Tabel 5.5 Hasil Confusion Matrix Data Uji 20 %

	<i>True Positive</i>	<i>False Negative</i>
<i>True Positive</i>	91	9
<i>False Negative</i>	0	100

$$\begin{aligned}
 \text{Akurasi} &= 91 + 100 / 91 + 9 + 0 + 100 \times 200 \\
 &= 191 / 200 \times 100 \\
 &= 95.5 \%
 \end{aligned}$$

Hasil pengujian ketiga menggunakan metode *support vector machine* pada data uji sebanyak 30% dengan jumlah data sebanyak 300 komentar, hasil akurasi sebesar 97.33 %. Hasil prediksi menggunakan *confosion matrix* dapat dilihat pada tabel berikut.

Tabel 5.6 Hasil Confusion Matrix Data Uji 30 %

	<i>True Positive</i>	<i>False Negative</i>
<i>True Positive</i>	144	6
<i>False Negative</i>	2	148

$$\begin{aligned}
 \text{Akurasi} &= 144 + 148 / 144 + 6 + 2 + 148 \times 300 \\
 &= 292 / 300 \times 100 \\
 &= 97.33 \%
 \end{aligned}$$

Pengukuran validasi juga dilakukan dengan *k-fold validation* agar data yang dihasilkan terbukti keakuratannya. Berikut akan ditampilkan hasil validasi akurasi dengan *k-fold*:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Hasil prediksi SVM dengan menggunakan data uji 10% dapat dilihat pada gambar dibawah.

K- 1
9300.0
K- 2
9300.0
K- 3
9300.0
K- 4
9300.0
K- 5
9300.0
K- 6
9300.0
K- 7
9300.0
K- 8
9300.0
K- 9
9300.0
K- 10
9300.0

Gambar 5.10 Implementasi *K-Fold* 10%

2. Hasil prediksi SVM dengan menggunakan data uji 20% dapat dilihat pada gambar dibawah.

K- 1
9550.0
K- 2
9550.0
K- 3
9550.0
K- 4
9550.0
K- 5
9550.0
K- 6
9550.0
K- 7
9550.0
K- 8
9550.0
K- 9
9550.0
K- 10
9550.0

Gambar 5.11 Implementasi *K-Fold* 20%

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

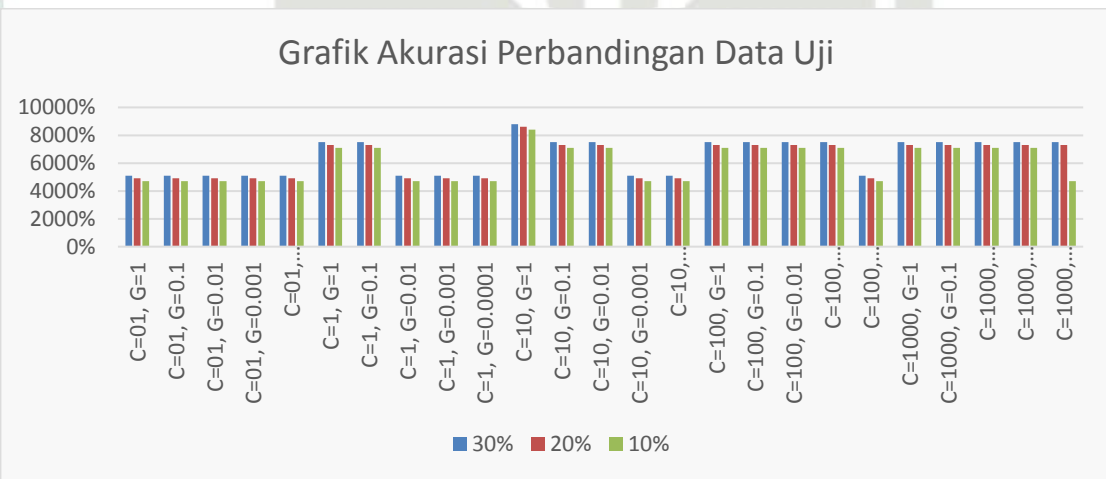
3. Hasil prediksi SVM dengan menggunakan data uji 20% dapat dilihat pada gambar dibawah.

```

K- 1
9733.33333333332
K- 2
9733.33333333332
K- 3
9733.33333333332
K- 4
9733.33333333332
K- 5
9733.33333333332
K- 6
9733.33333333332
K- 7
9733.33333333332
K- 8
9733.33333333332
K- 9
9733.33333333332
K- 10
9733.33333333332
    
```

Gambar 5.12 Implementasi K-Fold 30%

Pada percobaan menggunakan metode pembagian data dengan analisa *confusion matrix* dan *k-fold validation* di dapat hasil akurasi algoritma SVM terhadap klasifikasi sentimen pada video Ustad Abdul Somad, dimana pada pengujian pada data uji 30 % dengan jumlah 300 data komentar menghasilkan nilai akurasi tertinggi yaitu sebesar 97.33 % Perbandingan masing-masing pengujian data uji dapat dilihat pada gambar dibawah.



Gambar 5.13 Grafik Parameter Nilai (C, Gamma)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.3.1 Analisa Hasil Evaluasi *Confusion Matrix*

Untuk mengetahui keberhasilan dari tujuan evaluasi, akan dibandingkan hasil klasifikasi sentimen dari model pembelajaran yang dibangun sesuai dengan label kelas yang sudah ditentukan sebelumnya. Penelitian ini menggunakan 1000 dataset data komentar yang digunakan terdiri dari 70 % atau 700 data sebagai data latih dan 30 % atau 300 data sebagai data uji. Untuk evaluasi model pembelajaran klasifikasi sentimen menggunakan nilai parameter ($C = 10$ dan $\text{Gamma} = 1$). Berikut adalah tabel rincian data pada hasil pengujian:

Tabel 5.7 Rincian Pengujian Akurasi Tertinggi

Data Komentar	1000
Data Latih	700
Data Uji	300
True Positif	144
True Negatif	148
Akurasi	97.33 %

5.3.2 Kesimpulan Hasil Evaluasi

Penelitian ini menggunakan metode *support vector machine* dalam mengklasifikasikan data komentar terhadap video youtube Ustadz Abdul Somad yang menggunakan bahasa Indonesia. Dataset yang digunakan adalah sebanyak 1000 yang terbagi menjadi dua bagian yaitu data latih dan data uji. Dalam melakukan klasifikasi pada penelitian ini mendapatkan akurasi paling besar pada perbandingan data 70:30 dengan data latih sebanyak 700 dan data uji sebanyak 300 yaitu akurasi sebesar 97.33%.