

## **BAB 3**

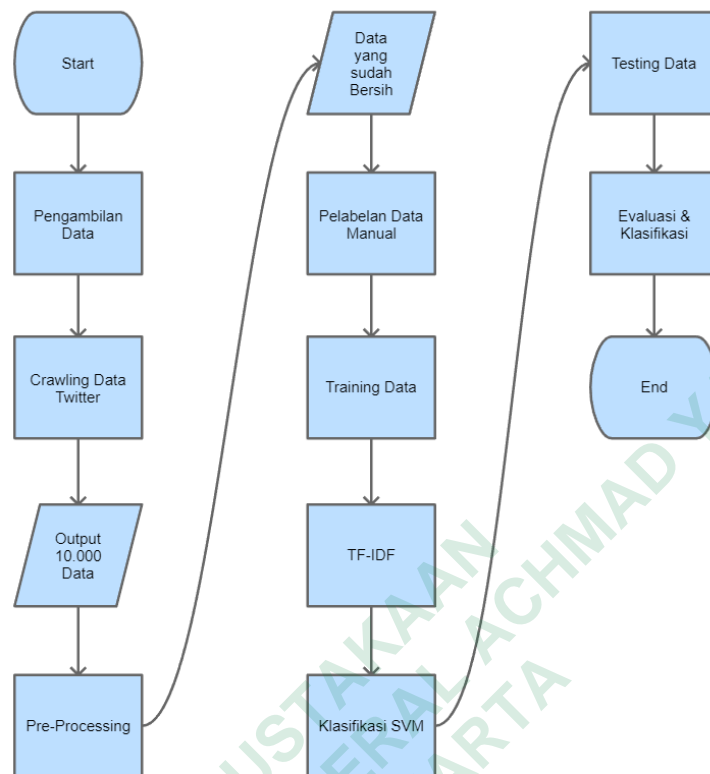
### **METODE PENELITIAN**

#### **3.1 BAHAN DAN ALAT PENELITIAN**

Bahan penelitian diambil melalui platform Twitter, mengambil beberapa data dan informasi dengan scrapping untuk pengambilan data, dimana aplikasi dapat mengambil beberapa tweet pada platform Twitter. Dengan beberapa kata kunci yang dapat digunakan untuk mencari data yang diperlukan untuk melakukan analisis, antara lain: “BPJS syarat SIM, BPJS Persyaratan SIM, BPJS SIM, BPJS Kesehatan SIM, #BPJS, #BPJSSIM, #SIMBPJS, #syaratSIM, fungsi BPJS SIM, BPJS Ketenaga Kerjaan SIM”. Proses pengambilan data dilakukan pada tanggal 10 Agustus 2022 - 13 November 2022, mendapatkan data sebanyak 112.668 data tweet pengambilan data pada tanggal tersebut agar mendapatkan data yang di inginkan dari mulai awal adanya regulasi tersebut.

Alat penelitian yang digunakan adalah komputer dengan spesifikasi cukup untuk menjalankan sistem operasi dan perangkat lunak pengembangan serta koneksi Internet. Sistem operasi dan beberapa program perangkat lunak yang digunakan dalam melakukan pengembangan aplikasi, sebagai berikut :

1. Sistem oprasi : Windows 10 Home Single Language 64-bit
2. Prosesor Intel Core i5 8265U @ 1.60GHz Kaby Lake-U/Y 14nm Technology, Ram 12GB, Storage 931GB
3. Bahasa Pemrograman Python 3.10.3
4. Anaconda 3 versi4.8.2 64bit
5. Excel 2016
6. Sublime text
7. Library SNS
8. Google Collab



**Gambar 3.1** Flowchart Alur Penelitian

Tahapan pada alur penelitian dapat dilihat pada Gambar 3.1 yaitu meliputi tahap Pengambilan Data, *Pre-Processing*, Pelabelan Data Manual, *Training Data*, TF-IDF, Klasifikasi SVM, *Testing Data*.

### 3.1.1 Pengambilan Data

Pada proses pertama yaitu pengambilan data pada Twitter, data yang diambil dengan memasukan beberapa kata kunci “BPJS syarat SIM, BPJS Persyaratan SIM, BPJS SIM, BPJS Kesehatan SIM, #BPJS, #BPJSSIM, #SIMBPJS, #syaratSIM, fungsi BPJS SIM, BJPS Ketenaga Kerjaan SIM” mulai dari tanggal 10 Agustus 2022 – 13 November 2022. Pengambilan data menggunakan *Snsrape* melalui Google Colab, lalu data yang sudah diambil dalam format XLSX. *Library* untuk pengambilan data dapat dilihat pada kode dibawah ini.

```
pip install snsrape
```

Kode diatas menampilkan kode *library* untuk melakukan pengambilan data tweet yang dibutuhkan, dengan cara awal yaitu menginstall *snsrape*. Apabila *library* sudah ter install selanjutnya tahap yang dilakukan untuk pemanggilan *library* dapat dilanjutkan seperti kode *library* dibawah.

```
import snsrape.modules.twitter as sntwitter
import pandas as pd
```

Diatas menampilkan pemanggilan *library* dengan melakukan *import* untuk digunakan pada tahapan selanjutnya. Kode yang digunakan untuk pengambilan data tweet pada Twitter dapat dilihat pada kode dibawah ini.

```
tweet_data = []
username = input('Enter your keyword: ')
number = int(input('How many tweets do you to scrape: '))
for i,tweet in enumerate(sntwitter.TwitterSearchScrapers('bpjs
sim since:2020-01-01 until:2022-11-13 lang:id').get_items()):
    if i > number:
        break
    tweet_data.append([tweet.date, tweet.id, tweet.content,
tweet.username])
df = pd.DataFrame(tweet_data, columns=['Date', 'Tweets',
'Content', 'Username'])
df.to_csv(f'{username}.csv', index=False, encoding='utf-8')
```

kode diatas merupakan fungsi untuk melakukan ekspor data tweet. Apabila setelah dijalankan maka akan muncul seperti Gambar 3.2

Enter your keyword:

**Gambar 3.2** Memasukan keyword yang digunakan

Pada gambar diatas digunakan untuk memberikan nama pada file data tweet yang ingin diambil dan ekspor kedalam csv. Setelah memasukan nama untuk file maka selanjutnya akan muncul seperti Gambar 3.3.

How many tweets do you to scrape:

**Gambar 3.3** Banyak Data yang di inginkan

Pada gambar 3.3 memiliki fungsi yaitu berapa banyak data Twitter yang ingin diambil dan diekspor ke dalam csv. Hasil dari data tweet yang berhasil di ekspor dapat dilihat pada Tabel 3.1.

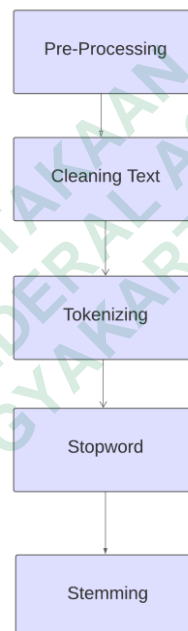
**Tabel 3.1** Data Tweet

No	Data Tweet
1.	Kepesertaan BPJS Kesehatan rencananya akan diwajibkan untuk calon jamaah haji dan umrah, permohonan SIM, STNK, dan Surat Keterangan Catatan Kepolisian (SKCK).
2.	Yang gaji gede bantu masyarakat penghasilan di bawah/ gaji rendah fine aja sih, yg jd masalah adalah memaksakan BPJS utk ngurus Sim, Perpanjangan Stnk dll ini yg ga bener. @BPJSKesehatanRI mestinya jg wajibkan pengusaha kaya wajibkan ikut BPJS biar membantu Negara.
3.	BPJS Kesehatan Jadi Syarat Urus SIM, SKCK dan STNK, Sukarno: Kebijakan Kurang Tepat
4.	Surat Izin Mengemudi alias SIM wajib dimiliki oleh pengendaraan kendaraan. Bagi Anda yang mau mengurus SIM baik bikin atau memperpanjangnya harus juga memiliki kartu peserta Badan Penyelenggaraan Jaminan Sosial (BPJS) Kesehatan sebagai syaratnya.
5.	@lantasresgresik Min terkait aturan terbaru mengenai syarat usulan SIM dan STNK yang harus punya BPJS Kesehatan, jika hanya punya BPJS Ketenagakerjaan apakah boleh?
6.	@exewinnie @BPJSKesehatanRI Ga. Sy br aja perpanjang sim ga ditanya bpjs samsek. Lancar joyo 25 menit di samsat jaktim
7.	#Paripurna36 Hillary B. @NasDem #Sulut: Terkait BPJS, sekarang apa-apa harus BPJS, kita ingin membuat SIM, jual beli rumah butuh BPJS, namun BPJS tidak menanggung biaya pelaporan kepada korban terkait kekerasan seksual. Memang kita DPR dengan fasilitas kesehatan premium.
8.	Bikers Kurang Setuju BPJS Kesehatan Jadi Syarat Bayar Pajak Motor dan Perpanjang SIM
9.	Pak @ListyoSigitP , untuk urus sim, harus kesehatan+psikologi yg tidak terkoneksi dengan bpjs kesehatan. Padahal polri adalah bagian dari pemerintah, seharusnya kesehatan+psikologi nya bisa menggunakan kartu BPJS. Mohon Pak @jokowi segera perbaiki hal ini.
10.	Jangan jadikan kesehatan masyarakat yang terkait dengan nyawa jadi permainan politik dan bisnis oligarki. Kejahatan yang tersamar dibalik pemaksaan kebijakan BPJS Kes untuk syarat administrasi pengurusan

hajat masyarakat (SIM, STNK, jual beli tanah, naik haji dll) <a href="https://t.co/8m21qdGLCt">https://t.co/8m21qdGLCt</a> ",Kimberley20101
--

Pada Tabel 3.1 bahwa hasil dari pengambilan data yang diekspor memiliki komponen-komponen yang perlu dihilangkan karena tidak diperlukan untuk melakukan analisis sentimen, beberapa komponen inilah yang nantinya akan dilakukan *preprocessing* data agar komponen tersebut dapat hilang dan mendapatkan hasil yang sesuai atau dibutuhkan dalam melakukan analisis sentimen.

### 3.1.2 Preprocessing Data



**Gambar 3.4** Flowchart Preprocessing

*Preprocessing* data adalah pengolahan data untuk menyiapkan data yang belum siap untuk dilakukan ketahapan selanjutnya. *Library* yang digunakan pada tahap ini berupa *pandas* untuk mengatur data mentah agar menjadi data yang terstruktur, *numpy* digunakan untuk mengolah data numerik, dan *nlTK* digunakan untuk membantu pemrosesan dari data teks pada data mentah.

```

import pandas as pd, numpy as np, nltk, string, re
from pandas import DataFrame
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
  
```

## 1. *Cleaning Text*

*Cleaning Text* merupakan penghapusan teks yang tidak digunakan seperti html, url, *emoji*, *mention*, *hashtag*, dan menghilangkan simbol atau beberapa huruf yang tidak perlu.

```
def cleaning_content(content):
    tab=content.replace('\t','').replace('\n','').replace('\',
    ' ')
    score = tab.replace('_', '')
    user = re.sub('@[A-Za-z0-9]+', '', score)
    link = re.sub(
        '((https?):(//)|(\.\.\/))+([\w\d:#@%/;$()~_?+\=\\.&](#!)?)*
        )+', '', user)
    url = re.sub(r'http\S+', '', link)
    punc = re.sub(r'^\w\s', '', url)
    rt = re.sub(r'RT[\s]+', '', punc)
    no = re.sub('[0-9]+', '', rt)
    slang = re.sub(r'\n', " ", no)
    reg = re.sub("b'", " ", slang)
    hashtag = re.sub('/#[\w_]+[ \t]*/', '', reg)
    result = hashtag.lower()
```

Kode diatas merupakan kode yang digunakan pada cleaning data untuk menghapus simbol-simbol yang tidak diperlukan. Hasil dari tahap *cleaning* data dapat dilihat pada Tabel 3.2.

**Tabel 3.2** Data *Cleaning Text*

No	Data <i>Tweet</i>
1.	Kepesertaan BPJS kesehatan rencananya akan diwajibkan untuk calon jamaah haji dan umrah permohonan sim stnk dan surat keterangan catatan kepolisian skck
2.	Seharusnya sekalian wajib bawa bpjs kesehatan ketenagakerjaan sim STNK dan bpkb tagihan listrik dan pbb xixixi
3.	Buat pak kapolri saya mau nanya nie pak kenapa pas urus sim untuk check kesehatancheck psikologi nya tdk bisa gunakan bpjs kesehatan cc pak cc.
4.	Klo bisa bpjs untuk yg mau saja biar tidak ada unsur paksaan jadi ben berkah tapi gk mungkin karna klo gak punya bpjs gak bisa bikan sim stnk
5.	BPJS kesehatan jadi syarat urus sim skck dan stnk sukarno kebijakan kurang tepat
6.	urus sim harus punya BPJS kesehatan akan secepatnya berlaku

7.	mengurus sim harus punya BPJS kesehatan kapan mulai berlaku ini kata korlantas polri
8.	surat izin mengemudi alias sim wajib dimiliki oleh pengendara kendaraan bagi anda yang mau mengurus sim baik bikin atau memperpanjangnya harus juga memiliki kartu peserta badan penyelenggaraan jaminan sosial bpjs kesehatan sebagai syaratnya
9.	Siap siap urus sim stnk dan bpkb harus menyertakan kartu BPJS kesehatan
10.	Bikers kurang setuju BPJS kesehatan jadi syarat bayar pajak motor dan perpanjang SIM

Pada Tabel 3.2 menampilkan data tweet yang sudah di *cleaning* data, sehingga mendapatkan data tweet yang bersih dan tidak mengandung html dan url, *emoji*, *hashtag*, dan *manton*.

## 2. *Tokenizing*

*Tokenizing* adalah proses untuk memecah kalimat-kalimat menjadi kata. Kode yang digunakan pada proses *tokenizing* dapat dilihat dibawah ini.

```
def clean_data(tweet)=
    for i in content:
        clean_data.append(cleaning_content(i))
    clean_content = pd.DataFrame()
    clean_content['Content'] = clean_data
    clean_content(data["Content"])
```

Hasil dari proses *Tokenizing* dapat dilihat pada Tabel 3.3

**Tabel 3.3** Hasil *Tokenizing*

<p>"['perpanjang', 'sim', 'stnk', 'hingga', 'skck', 'wajib', 'punya', 'bpjs', 'kesehatan']",          "['aturan', 'kebijakan', 'semakin', 'tdk', 'jelas', 'nglantur', 'jual', 'beli', 'bikin', 'sim',          'hrs', 'pny', 'bpjs', 'bpjs', 'itu', 'hubunganya', 'kesehatan', 'sakit', 'skrng',          'pengendara', 'di', 'larang', 'pke', 'sendal', 'jepit']",          "['sulitnya', 'mencoba',          'perpanjang', 'sim', 'cara', 'online', 'lewat', 'aplikasi', 'banyak', 'kendala', 'klik',          'link', 'untuk', 'baca', 'ngga', 'cuma', 'sim', 'doang', 'kali', 'paspor', 'bpjs', 'kesehatan',          'dll', 'kayak', 'gini', 'mau', 'pake', 'teknologi', 'mobil', 'listrik', 'milih', 'server', 'aja',          'cupu']".</p>
--

### 3. *Stopword*

*Stopword* adalah proses yang digunakan untuk menghilangkan kata yang tidak memiliki arti. Kode yang digunakan pada tahap *stopword* dapat dilihat dibawah ini.

```
factory = StopWordRemoverFactory()
more_stopword = []
stopword = factory.create_stop_word_remover()
stopwords = factory.get_stop_words()+more_stopword
def removeStopWords(content):
    clean_word_list = [word for word in content.split() if
    word not in stopwords]
    return clean_word_list

stopwords_content = lower_tweet.apply(removeStopWords)

print(stopwords_content)
```

Hasil dari proses *Stopword* dapat dilihat pada Tabel 3.4.

**Tabel 3.4** Hasil *Library* Sastrawi

Hasil kata dari Library Sastrawi
"['kepesertaan', 'bpjs', 'kesehatan', 'rencananya', 'diwajibkan', 'calon', 'jamaah', 'haji', 'umrah', 'permohonan', 'sim', 'stnk', 'surat', 'keterangan', 'catatan', 'kepolisian', 'skck']"['sekalian', 'wajib', 'bawa', 'bpjs', 'kesehatan', 'ketenagakerjaan', 'sim', 'stnk', 'bpkb', 'tagihan', 'listrik', 'pbb', 'xixixi']"
"['berapa', 'ribu', 'orang', 'terpaksa', 'rubah', 'data', 'alamat', 'ktpstnsim', 'nya', 'karyawan', 'rubah', 'data', 'sistem', 'lapor', 'bpjs', 'ketenagakerjaan', 'kesehatan', 'hebat']"

Tabel 3.4 diatas merupakan hasil dari tahap *Stopword*, yaitu data yang tidak memiliki makna yang nantinya akan dihilangkan. Kode yang digunakan untuk menghilangkan kata yang tidak memiliki makna, dapat dilihat dibawah ini.

### 4. *Stemming*

*Stemming* merupakan proses untuk merubah kata menjadi kata dasar. Kode yang digunakan pada proses *Stemming* dapat dilihat dibawah ini.

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()
def stemmed_wrapper(term):
    return stemmer.stem(term)
term_dict = {}
for document in stopwords_content:
    for term in document:
```



```

        if term not in term_dict:
            term_dict[term] = " "
    print(len(term_dict))
    print("-----")
    for term in term_dict:
        term_dict[term] = stemmed_wrapper(term)
        print(term,":", term_dict[term])
    print(term_dict)
    print("-----")
    def get_stemmed_term(document):
        return [term_dict[term] for term in document]
    stem_content = stopwords_content.apply(get_stemmed_term)
    print(stem_content)

```

hasil dari proses *Stemming* dapat dilihat pada Tabel 3.5

**Tabel 3.5** Hasil *Stemming* Data

Hasil kata dari Library Sastrawi
['serta bpjs sehat rencana wajib calon jamaah haji umrah mohon sim stnk surat terang catat polisi skck'], ['gapake soal dulu sempet wacana ngurus sim amp stnk surat tanah haji umroh pake bpjs sehat'], ['sulit coba panjang sim cara online lewat aplikasi banyak kendala klik link baca ngga cuma sim doang kali paspor bpjs sehat kayak gin mau pake teknologi mobil listrik milih server aja cupu'], ['jangan urus stnk sim lama pake bpjs sehat yg gak huhungannya jangan bayar semua pajak aja masih wajib bpjs'], ['laku bpjs sehat jadi syarat wajib bikin sim stnk urus sim stnk wajib daftar anggota aktif bpjs sehat wacana sudah sejak maret silam kapan laku'], ['surat izin kemudi alias sim wajib milik kendra kendra anda mau urus sim baik bikin panjang juga milik kartu serta badan selenggara jamin sosial bpjs sehat syarat'], ['konsep baik eksekusi bagaimana juta tahun asuransi sehat tradisional udh bagus banget lo benefitnya kaum tengah seperti bpjs cuma jadi syarat duduk ktp sim']

## 5. Normalisasi

Normalisasi adalah proses yang digunakan untuk mengubah menjadi kalimat yang baku. Kode yang digunakan pada proses normalisasi dapat dilihat dibawah ini.

```

normalizad_word_dict = {}
for index, row in normalizad_word.iterrows():
    if row[0] not in normalizad_word_dict:
        normalizad_word_dict[row[0]] = row[1]
def normalized_term(document):
    return [normalizad_word_dict[term] if term in
            normalizad_word_dict else term for term in document]
normal_content = stem_content.apply(normalized_term).str.join(" ")
print(normal_content)

```

Pada kode diatas merupakan proses normalisasi dilakukan perubahan yang tidak baku menjadi baku sesuai dengan data yang ada. Contoh kata yang diproses pada tahap normalisasi dapat dilihat pada Tabel 3.6.

**Tabel 3.6** Daftar Kata Normalisasi

No	Sebelum	Sesudah
1.	tak	Tidak
2.	lancaaar	Lancer
3.	ngnsep	Konsep Saja
4.	tau	Tahu
5.	gajadi	Tidak jadi
6.	Tp	Tetapi
7.	yg	Yang
8.	tdk	Tidak
9.	gw	Saya
10.	pake	Pakai

Hasil dari *Pre-Processing* data dapat dilihat pada Tabel hasil *Pre-Processing* Data pada Tabel 3.7. data hasil *Pre-Processing* digunakan untuk melakukan Pelabelan manual.

Tabel 3.7 Hasil *Pre-Processing*

No	Data Tweet hasil preprocessing
1.	Kepesertaan BPJS kesehatan rencananya akan diwajibkan untuk calon jamaah haji dan umrah permohonan sim stnk dan surat keterangan catatan kepolisian skck
2.	Seharusnya sekalian wajib bawa bpjs kesehatan ketenagakerjaan sim STNK dan bpkb tagihan listrik dan pbb xixixi
3.	Buat pak kapolri saya mau nanya nie pak kenapa pas urus sim untuk check kesehatancheck psikologi nya tdk bisa gunakan bpjs kesehatan cc pak cc.
4.	Klo bisa bpjs untuk yg mau saja biar tidak ada unsur paksaan jadi ben berkah tapi gk mungkin karna klo gak punya bpjs gak bisa bikin sim stnk
5.	BPJS kesehatan jadi syarat urus sim skck dan stnk sukarno kebijakan kurang tepat
6.	urus sim harus punya BPJS kesehatan akan secepatnya berlaku
7.	mengurus sim harus punya BPJS kesehatan kapan mulai berlaku ini kata korlantas polri
8.	surat izin mengemudi alias sim wajib dimiliki oleh pengendaraan kendaraan bagi anda yang mau mengurus sim baik bikin atau memperpanjangnya harus juga memiliki kartu peserta badan penyelenggaraan jaminan sosial bpjs kesehatan sebagai syaratnya
9.	Siap siap urus sim stnk dan bpkb harus menyertakan kartu BPJS kesehatan
10.	Bikers kurang setuju BPJS kesehatan jadi syarat bayar pajak motor dan perpanjang SIM

### 3.1.3 Pelabelan Manual

Tahap pelabelan ini dilakukan dengan melabeli kalimat yang berada pada data tweet secara manual, untuk dianalisa lebih lanjut mengenai kalimat *positif*, dan *negative*. Pelabelan manual ini dari data yang dihasilkan dari tahap *Pre-Processing* terlebih dahulu. Data yang didapat dari *Pre-Processing* sebanyak 1470 data tweet dan dilakukan pelabelan manual sebanyak 1470 data tweet dan dilabeli, 735 data tweet berlabel Positif, dan 735 data tweet berlabel Negatif. Contoh tweet yang akan diberi label secara manual dapat dilihat pada Tabel 3.8.

Pelabelan	
Show 10 entries	Search: <input type="text"/>
Content	Centiment
ada aturan baru nih mau ngurus sim dan stnk serta calon jemaah umrah dan haji syaratnya harus jadi peserta bpjs kesehatan bgmn implementasi kedepannya apa saja catatan terkait implementasi inpres ini whatstrending podcastindonesia	OPositif ONetral ONegatif
ada berita buat sim dan stnk wajib ada kartu bpjs apakah juga berlaku untuk perpanjangan sim dan stnk min	OPositif ONetral ONegatif
ada bpjs kelas yg disubsidi pelayanan sama orang yg berhubungan dgn sim stnk bpkb atau jual beli tanah pastilah orang yg cukup mampu utk ikut bpjs sekedar kelas hanya utk syarat utk yg miskin ada jkn yg iurannya ditanggung pemerintah	OPositif ONetral ONegatif
ada lagi nih yg baru warga wajib punya kartu bpjs kesehatan urus sim stnk hingga naik haji	OPositif ONetral ONegatif
ada rencana utk tdk bayar iuran jkn krn tdk rela iurannya diselewengkan ke ikn eehhh langsung keluar berita ini urus sim hingga daftar umrah wajib jadi peserta bpjs kesehatan	OPositif ONetral ONegatif
ada yg sangat peduli kesehatan kita yang ingin membeli tanah dengan mewajibkan bpjs jadi kalau kita ingin berkebun di tanah itu dan tidak senoaia kena cangkok tenano saia langsung rebhat gratis dennan bnis baik sekaleee kenedulian vn sama saat kita urus sim amp stnk	OPositif ONetral ONegatif

**Gambar 3.5** Pelabelan manual

**Tabel 3.8** Hasil Pelabelan Manual

No	Tweet	Kelas	Nilai
1	Kepesertaan BPJS Kesehatan rencananya akan diwajibkan untuk calon jamaah haji dan umrah, permohonan SIM, STNK, dan Surat Keterangan Catatan Kepolisian (SKCK).	Positif	1
2	Presiden Joko Widodo sudah menekan Inpres Nomor 1 Tahun 2022 tentang Optimalisasi Pelaksanaan Program Jaminan Kesehatan Nasional. Dalam aturan itu, masyarakat yang hendak mengurus SIM hingga STNK harus terdaftar dalam kepesertaan aktif BPJS Kesehatan	Positif	1
3	Saya hidup di era serba membingungkan. Bpjs kesehatan yg fungsi utamanya untuk kesehatan digunakan untuk syarat administrasi stnk, sim, skck dll. Tes Covid-19 yg fungsi utamanya untuk screening kesehatan digunakan pula untuk syarat berpergian.	Negative	-1
4	Syarat BPJS Kesehatan Dinilai Bikin Warga Malas Urus SIM dan STNK	Negative	-1

Dari Tabel 3.8 diatas terdapat hasil bahwa label positif diberi kelas dengan nilai 1, dan label negatif diberi kelas dengan nilai -1. Dan pelabelan manual ini dilakukan didalam *website*, apabila belum selesai melakukan pelabelan maka dapat dilakukan *draft* dahulu dan jika sudah ter labeli keseluruhan maka dapat dilakukan *submit* untuk penyelesaian pelabelan manual., dan menghasilkan file pelabelan

dengan format csv. Pelabelan manual ini dilakukan agar memberikan hasil tentang nilai sentimen yang memiliki kelas positif, atau negatif yang nantinya akan dihitung nilai akurasinya.

### 3.1.4 Training Data

*Training* data adalah sebuah proses training pada data tweet dengan menggunakan metode SVM. Dimana pada tahap ini akan dilakukan pelabelan manual terlebih dahulu setelah melakukan *Pre-Processing* Data, dan menghasilkan 1470 data tweet yang sudah terlabeli dengan rincian data 735 berlabel Positif dan 735 berlabel Negatif. Lalu dilanjutkan dengan proses *training* untuk menghasilkan model klasifikasi yang digunakan untuk *Testing* Data nanti. Kode *training* data bisa dilihat dibawah ini:

```
@app.route('/training', methods=["POST", "GET"])
def upload():
    try:
        if request.method == 'POST':
            # upload file flask
            upload_file = request.files['file']
            # Extracting uploaded data file name
            data_filename = secure_filename(upload_file.filename)
            # flask upload file to database (defined uploaded folder in
static path)
            upload_file.save(os.path.join(application.config['UPLOAD_FOL
DER'], data_filename))
            # Storing uploaded file path in flask session
            session['uploaded_data_file_path'] =
os.path.join(application.config['UPLOAD_FOLDER'], data_filename)
            return redirect(url_for('index'))

        # read xlsx file in python flask (reading uploaded xlsx file
from uploaded server location)
        uploaded_df =
pd.read_excel(os.path.join(application.config['UPLOAD_FOLDER'],
"data_training.xlsx"))
        uploaded_df = pd.DataFrame(uploaded_df)
        print(uploaded_df)
        # pandas dataframe to html Tabel flask
        uploaded_df_html = uploaded_df.values
        p = uploaded_df[uploaded_df['label'] ==
'positif']['kelas'].count()
        n = uploaded_df[uploaded_df['label'] ==
'negatif']['kelas'].count()
        return render_template('upload_excel.html',
data=uploaded_df_html, p=p, n=n, o=o)
    except Exception as e:
        return render_template('upload_excel.html')
```

Kode yang digunakan untuk pemanggilan data yang sudah di *training* manual dapat dilihat dibawah ini.

```
df = pd.read_excel ('data_training.xlsx')
df=pd.DataFrame(df)
df=df.dropna()
df
```

Pada kode tersebut menampilkan kode yang difungsikan untuk memanggil data *training* yang sudah melalui proses *Pre-Processing* dan pelabelan data manual. Hasil dari data *training* ditampilkan dan dapat dilihat pada Gambar 3.6.

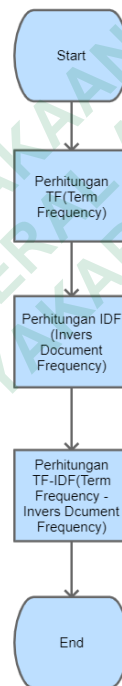
	Date	Content	Username	Label	Kelas
1	2022-08-12 15:16:37+00:00	@worksfess Kalau minta KTP/KK/SIM skip ya. It...	ivanasilo	N	-1
7	2022-08-10 11:10:11+00:00	BPJS jadi syarat perpanjangan atau buat baru S...	balikpapanti	O	0
8	2022-08-10 10:10:54+00:00	@crestfallen89 iyaa ada sim, ktp, stnk, bpjs, ...	bbycomeon	O	0
11	2022-08-09 09:25:14+00:00	@zarazettirazr Paling males kalau ngurus KTP, ...	tatakhoiriyah	N	-1
14	2022-08-06 04:17:16+00:00	SUMPAH YA AKU TUH jarang bawa dompet sampe gat...	nadachalamet	O	0
...	...	...	...	...	...
8962	2022-02-21 09:57:08+00:00	Setujukah Anda dengan ketentuan ini?\n\nTak ha...	metro tv	O	0
8963	2022-02-21 09:50:23+00:00	boro boro kartu sembako murah 😞 Skarang malah...	mwap	N	-1
8965	2022-02-21 09:22:24+00:00	@tembang_kangen Delo neh Kartu Indonesia Sehat...	msnugroho	O	0
8966	2022-02-21 09:13:13+00:00	@_ekokuntadhi Nggak ada relevansinya antara ka...	ardhy	N	-1
8969	2022-02-21 08:50:45+00:00	Ketua DPD RI, AA LaNyalla Mahmud Mattalitti, m...	adibaprayoga	N	-1

**Gambar 3.6** Hasil pemanggilan data *training*

Pada Gambar 3.6 adalah hasil pemanggilan data *training* yang telah di labeli secara manual. Contoh dokumen yang nantinya akan digunakan untuk perhitungan TF-IDF dapat dilihat pada Tabel 3.8.

Tabel 3.8 Term/Kata

(D) Dokumen	Kalimat
D1	JAKE SIM MAJU LOOOO GUE GATAKUT GUE PUNYA BPJS
D2	Makin lawak aja perpanjang sim stnk harus ada bpjs kesehatan (yang sebenarnya masih gak nemu korelasinya apa)
D3	Kalau (Pancasila) ada, masak iuran kesehatan (BPJS) menjadikan masyarakat sampai kejet-kejet, gak boleh perpanjang SIM, paspor. Di mana Pancasila?"
D4	Perpanjang SIM harus bawa kartu BPJS Kesehatan, pengemudi ojol ketar-ketir: Kami tertindas!



Gambar 3.7 Alur Perhitungan TF-IDF

Pada Tabel 3.8 *term* atau kata memiliki 4 dokumen yang ter labeli D1 yang merupakan dokumen ke-1, D2 berisikan dokumen ke-2, D3 merupakan dokumen ke-3, dan D4 berisikan dokumen ke-4. Dokumen yang sudah ada, yang berisikan kalimat dari tweet yang sudah diproses akan digunakan untuk melakukan proses perhitungan IDF secara manual, dan perhitungan dari TF dapat dilihat pada Tabel 3.9.

Tabel 3.9 Perhitungan TF

Term	D1	D2	D3	D4	Df
jake	1				1
sim	1	1	1	1	4
maju	1				1
lo	1				1
gue	1				1
gatakut	1				1
punya	1				1
bpjs	1	1	1	1	4
makin	1				1
lawak	1				1
aja	1				1
perpanjang		1	1	1	3
stnk		1			1
harus		1		1	2
ada		1	1		2
kesehatan		1	1	1	3
yang		1			1
sebenarnya		1			1
masih		1			1
gak		1	1		2
nemu		1			1
korelasinya		1			1
apa		1			1
kalau			1		1
pancasila			2		1
masak			1		1
iuran			1		1
menjadikan			1		1
masyarakat			1		1
kejut-kejut			1		1
boleh			1		1
paspor			1		1
dimana			1		1
bawa				1	1
kartu				1	1
pengemudi				1	1
ojol				1	1
ketar-ketir				1	1
kami				1	1
tertindas				1	1



Pada Tabel 3.9 perhitungan TF diatas data yang digunakan memiliki persamaan (1) pada setiap kata untuk dilakukan perhitungan. Perhitungan IDF dapat dilihat pada Tabel 3.10.

**Tabel 3.10** Perhitungan IDF

Term	df	idf	idf(D=4)
jake	1	1	0,60205
sim	4	0,25	0
maju	1	1	0,60205
lo	1	1	0,60205
gue	1	1	0,60205
gatakut	1	1	0,60205
punya	1	1	0,60205
bpjs	4	0,25	0
makin	1	1	0,60205
lawak	1	1	0,60205
aja	1	1	0,60205
perpanjang	3	0,3333	0,12493
stnk	1	1	0,60205
harus	2	0,5	0,30102
ada	2	0,5	0,30102
kesehatan	3	0,3333	0,12493
yang	1	1	0,60205
sebenarnya	1	1	0,60205
masih	1	1	0,60205
gak	2	0,5	0,30102
nemu	1	1	0,60205
korelasinya	1	1	0,60205
apa	1	1	0,60205
kalau	1	1	0,60205
pancasila	1	1	0,60205
masak	1	1	0,60205
iuran	1	1	0,60205
menjadikan	1	1	0,60205
masyarakat	1	1	0,60205
kejet-kejet	1	1	0,60205
boleh	1	1	0,60205
paspor	1	1	0,60205
bawa	1	1	0,60205
kartu	1	1	0,60205
pengemudi	1	1	0,60205
ojol	1	1	0,60205

ketar-ketir	1	1	0,60205
kami	1	1	0,60205
tertindas	1	1	0,60205

Pada Tabel 3.9 menunjukkan dari perhitungan TF manual dengan persamaan 1 pada setiap kata dan pada Tabel 3.10 menunjukkan perhitungan IDF manual dengan rumus 1 sebagai persamaanya, setelah selesai dilakukan perhitungan tf dan idf secara manual, maka selanjutnya akan dilakukan perhitungan TF-IDF secara manual dapat dilihat pada. Tabel 3.11 Perhitungan TF-IDF secara Manual

**Tabel 3.11** Perhitungan TF-IDF Manual

Term	D1	D2	D3	D4
jake	0,60205			
sim	0	0	0	0
maju	0,60205			
lo	0,60205			
gue	0,60205			
gatakut	0,60205			
punya	0,60205			
bpjs	0	0	0	0
makin	0,60205			
lawak	0,60205			
aja	0,60205			
perpanjang		0,12493	0,12493	0,12493
stnk		0,60205		
harus		0,30102		0,30102
ada		0,30102	0,30102	
kesehatan		0,12493	0,12493	0,12493
yang		0,60205		
sebenarnya		0,60205		
masih		0,60205		
gak		0,30102	0,30102	
nemu		0,60205		
korelasinya		0,60205		
apa		0,60205		
kalau			0,60205	
pancasila			0,30102	
masak			0,60205	
iuran			0,60205	
menjadikan			0,60205	

masyarakat			0,60205	
kejet-kejet			0,60205	
boleh			0,60205	
paspor			0,60205	
bawa				0,60205
kartu				0,60205
pengemudi				0,60205
ojol				0,60205
ketar-ketir				0,60205
kami				0,60205
tertindas				0,60205

Perhitungan TF-IDF dilakukan agar dapat menghasilkan perhitungan yang dapat dibobotkan pada setiap kata dan setiap dokumen yang digunakan secara otomatis pada google colab. Kode perhitungan TF-IDF tersebut dapat dilihat sebagai berikut.

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
d1 = "JAKE SIM MAJU LOOOO GUE GATAKUT GUE PUNYA BPJS"
```

```
d2 = "Makin lawak aja perpanjang sim stnk harus ada bpjs kesehatan(yang sebenarnya masih gak nemu korelasinya apa)"
```

```
d3 = "Kalau (Pancasila) ada, masak iuran kesehatan (BPJS) menjadikan masyarakat sampai kejet-kejet, gak boleh perpanjang SIM, paspor.Di mana Pancasila?"
```

```
d4 = "Perpanjang SIM harus bawa kartu BPJS Kesehatan, pengemudi ojol ketar-ketir:Kami tertindas!"
```

```
vect = TfidfVectorizer()
X = vect.fit_transform([d1, d2, d3, d4])
X.toarray()
```

```
[(yang 0.10369516947304253),(perpanjang 0.3110855084191276),
(gak 0.20739033894608505),(harus 0.20739033894608505),(bpjs 0.
41256849850351734),(masyarakat 0.10314212462587934),
(tertindas 0.10314212462587934),(pancasila 0.
20628424925175867),(masak 0.10314212462587934)]
```

**Gambar 3.8** Hasil perhitungan TF-IDF otomatis

Pada Gambar 3.8 merupakan hasil dari perhitungan TF-IDF yang dilakukan perhitungan otomatis pada Google Colab.

Kode perhitungan *confusion matrix* dapat dilihat dibawah ini.

```
from sklearn.model_selection import ShuffleSplit
```

```

from sklearn.metrics import accuracy_score, f1_score, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
X = df.Content
y = df.kelas
ss = ShuffleSplit(n_splits=10, test_size=0.2)
sm = SMOTE()
accs = []
f1s = []
cms = []
for train_index, test_index in ss.split(X):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]
    X_train_vect = vect.fit_transform(X_train)
    X_test_vect = vect.transform(X_test)
    X_train_res, y_train_res = sm.fit_resample(X_train_vect, y_train)
    SVM.fit(X_train_res, y_train_res)
    y_pred = SVM.predict(X_test_vect)
    accs.append(accuracy_score(y_test, y_pred))
    f1s.append(f1_score(y_test, y_pred, average='weighted'))
    cms.append(confusion_matrix(y_test, y_pred))
print("\nAverage accuracy across folds: {:.2f}%".format(sum(accs) / len(
accs) * 100))
print("\nAverage F1 score across folds: {:.2f}%".format(sum(f1s) / len(f
1s) * 100))
print("\nAverage Confusion Matrix across folds: \n {}".format(sum(cms) /
len(cms)))
y_pred

```

Pada kode diatas tersebut digunakan untuk perhitungan *Confusion Matrix* pada data *training*, Perhitungan tersebut dapat menghasilkan informasi pembandingan dari klasifikasi yang telah dilakukan, sehingga akan mengetahui hasil dari kemampuan pada model yang telah dibuat.

```

import os
import pickle
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder
X = training_data.Content
y = training_data.kelas

y = y.astype(str)
Encoder = LabelEncoder()
Train_y = Encoder.fit_transform(y)
Tfidf_vect = TfidfVectorizer()
Tfidf_vect.fit(X)
Train_X_Tfidf = Tfidf_vect.transform(X)
# Classifier - Algorithm - SVM

```

```
SVM = SVC(C=1.0, kernel='linear', degree=3, gamma='auto')
Content_pickle = SVM.fit(Train_X_Tfidf, Train_y)
# Simpan ke file pickle
with open('SVM_classifier.pickle', 'wb') as file:
pickle.dump(Content_pickle, file)
```

Pada kode tersebut digunakan untuk membuat dan menyimpan file dengan format pickel, yang dimana digunakan untuk membaca dan menyimpan data dari sebuah file dan akan digunakan untuk proses klasifikasi SVM.

### 3.1.5 Metode *Support Vector Machine*

Melakukan klasifikasi menggunakan metode SVM ini, pada proses sebelumnya sudah melakukan proses data yang dimana menghasilkan file yang memiliki format pickel, file tersebut digunakan untuk melakukan perhitungan *k-fold* pengujian data. Untuk pemanggilan file tersebut kode yang digunakan terdapat dibawah ini.

```
model = open('SVM_classifier.pickle', 'rb')
svm_classifier = pickle.load(model)
```

Dan Kode dibawah merupakan kode yang berfungsi untuk menghitung mengujian dari *k-fold Accuracy* dan *f-measure* pada data yang digunakan.

```
fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True, figsize=(16,9))
acc_scores = [round(a * 100, 1) for a in accs]
f1_scores = [round(f * 100, 2) for f in f1s]
x1 = np.arange(len(acc_scores))
x2 = np.arange(len(f1_scores))
ax1.bar(x1, acc_scores)
ax2.bar(x2, f1_scores, color='#559ebf')
# Place values on top of bars
for i, v in enumerate(list(zip(acc_scores, f1_scores))):
    ax1.text(i - 0.25, v[0] + 2, str(v[0]) + '%')
    ax2.text(i - 0.25, v[1] + 2, str(v[1]))
ax1.set_ylabel('Accuracy (%)')
ax1.set_title('SVM')
ax1.set_ylim([0, 100])
ax2.set_ylabel('F1 Score')
ax2.set_xlabel('Runs')
ax2.set_ylim([0, 100])
sns.despine(bottom=True, left=True)
plt.show()
```

Hasil dari menghitung pengujian *k-fold Accuracy* dan *f-measure* dapat dilihat pada Gambar 3.9.



**Gambar 3.9** Grafik k-fold Accuracy dan *f-measure*

Pada gambar 3.9 diatas menampilkan grafik dari nilai *accuracy* dan *f-measure* yang sudah dilakukan beberapa kali pengujian sampai mendapatkan hasil tertinggi, dan mendapatkan hasil yang diinginkan. Pengujian tertinggi mendapatkan *Accuracy* 72.1% dan *f-measure* 72.1%.

### 3.1.6 Testing

Pada proses *testing* data disini menggunakan hasil klasifikasi dari *training* data yang sudah dibuat, menggunakan 630 data tweet yang digunakan untuk sebagai uji coba.

```
data_testing = pd.read_excel(os.path.join(app.config['UPLOAD_FOLDER'],
"data_testing.xlsx"))
```

Pada kode diatas merupakan kode yang digunakan untuk pemanggilan data *testing* yang telah disimpan sebelumnya dalam bentuk file excel. Hasil tampilan data *testing* dapat dilihat pada Gambar 4.5.

Username	Contents	Label	Kelas
aabdillahth	@CNNIndonesia Syarat untuk buat/perpanjang SIM dan STNK sangat bertolak belakang dg ketentuan BPJS Kesehatan <a href="https://t.co/Tc3clik9IL">https://t.co/Tc3clik9IL</a>	N	-1.0
aapasajaa	Pertama dan terakhir pake @linkaja kalo sampe besok ttp dipersulit pembukaan blokir an Pernah hilang kartu bpjs, dm aja lgsg minta nomer ktp alamat dll udah cukup lgsg dikasih tau nomer peserta nya berapa. Hla iki LINKAJA, sampe suruh kirim foto KTP, KK, SIM	N	-1.0
aapasajaa	Pertama dan terakhir pake @linkaja kalo sampe besok ttp dipersulit pembukaan blokir an Pernah hilang kartu bpjs, dm aja lgsg minta nomer ktp alamat dll udah cukup lgsg dikasih tau nomer peserta nya berapa. Hla iki LINKAJA, sampe suruh kirim foto KTP, KK, SIM	N	-1.0
abadikinicom	Pembuatan SIM dan STNK Mulai 1 Maret 2022 Wajib Melampirkan BPJS Kesehatan <a href="https://t.co/qBPpBL10js">https://t.co/qBPpBL10js</a>	P	1.0
abang faey	Warga yang ingin membuat Surat Izin Mengemudi (SIM), STNK, melaksanakan ibadah Haji atau Umrah, bahkan jual beli tanah harus memiliki kartu BPJS Kesehatan sbg salah satu syarat. Yodah yuk kita boikot pembayaran pajak kendaraan dan tanah biar kolap sektor pajak, biar modyar.	O	0.0
abang faey	Warga yang ingin membuat Surat Izin Mengemudi (SIM), STNK, melaksanakan ibadah Haji atau Umrah, bahkan jual beli tanah harus memiliki kartu BPJS Kesehatan sbg salah satu syarat. Yodah yuk kita boikot pembayaran pajak kendaraan dan tanah biar kolap sektor pajak, biar modyar.	P	1.0
abdrossid	@AnakLolina2 @jokowi @BPJSKesehatanRI Namanya mau bayar stnk ,sim berarti dia punya motormobil pertanda dia orang mampu bro	O	0.0
abdulja	BPJS kesehatan syarat wajib untuk: 1. Membuat SIM, STNK dan SKCK 2.jual beli tanah 3.yarat pendaftaran umroh dan haji... BPJS kesehatan ngaripuheun masyarakat...	O	0.0
abdulja	BPJS kesehatan syarat wajib untuk: 1. Membuat SIM, STNK dan SKCK 2.jual beli tanah 3.yarat pendaftaran umroh dan haji... BPJS kesehatan ngaripuheun masyarakat...	O	0.0
abdulhoiri	Nanti lama kelamaan syarat berkendara juga wajib punya BPJS, jadi ketika ada razia yg lazimnya diperiksa STNK, SIM & KTP skrg tmbah 1 lagi yaitu BPJS, jika tidak punya akan ditilang dan disanksi... Negeraku oh negaraku... <a href="https://t.co/EMORCISPL">https://t.co/EMORCISPL</a>	N	-1.0

**Gambar 3.10** Hasil pemanggilan data *Testing*

Pada gambar 3.10 menampilkan hasil dari pemanggilan testing data yang berisikan data tweet yang digunakan, berisikan label Positif, Negatif dan kelas pada tiap tweetnya. Data *testing* tersebut akan digunakan untuk perhitungan *confusion matrix* data *testing* pada proses selanjutnya. Kode perhitungan dari *confusion matrix* dapat dilihat pada kode dibawah ini.

```
print(classification_report(y_test, y_pred))
print("\nAverage accuracy across folds: {:.2f}%".format(sum(accs) /
len(accs) * 100))
print("\nAverage F1 score across folds: {:.2f}%".format(sum(f1s) /
len(f1s) * 100))
print("\nAverage Confusion Matrix across folds: \n {}".format(sum(cms) /
len(cms)))
Average accuracy across folds: 67.11%
Average F1 score across folds: 67.09%
Average Confusion Matrix across folds:
[[102.5  45.4]
 [ 51.3  94.8]]
```

Pada kode diatas menghasilkan data *accuracy* sebesar 67.11% dan *f-measure* sebesar 67.09%. Pada perhitungan *Confusion Matrix* dari data *testing* dengan hasil TP sebesar 51.3, FN sebesar 94.8, FP sebesar 102.5, TN sebesar 94.8