

BAB 3

METODE PENELITIAN

Penelitian ini menganalisis sentimen negatif dan positif dari data Twitter tentang Ibu Kota Baru Nusantara yang akan pindah pada tahun 2024 dengan menggunakan metode *support vector machine* (SVM). Penelitian ini membutuhkan data *tweet* yang berkaitan dengan pendapat tentang pemindahan Ibu Kota Negara (IKN), yang kemudian akan dijalankan pengolahan data yaitu *preprocessing* untuk mendapatkan hasil yang diinginkan. Informasi atau data dalam hal ini akan digunakan untuk mengukur opini publik baik dari sudut pandang positif maupun negatif sehingga untuk menentukan tingkat kesetujuan oleh media tentang ibu kota negara yang akan pindah pada tahun 2024.

3.1 BAHAN DAN ALAT PENELITIAN

Bahan-bahan penelitian yang mungkin dibutuhkan untuk menganalisis sentimen opini publik tentang perpindahan ibu kota negara pada media sosial Twitter antara lain:

1. Data Twitter: Data Twitter yang terkait dengan topik perpindahan ibu kota negara. Data ini dapat diperoleh menggunakan teknik *scrapping* dari situs Twitter berjumlah 3000 data. Menggunakan kata kunci “Perpindahan Ibu Kota dan Ibu Kota Nusantara”.
2. Literatur terkait: Literatur terkait dapat membantu dalam memahami konteks dan isu-isu terkait dengan perpindahan ibu kota negara. Literatur terkait juga dapat digunakan untuk memperkaya analisis sentimen pada *tweet* yang dikaitkan dengan topik perpindahan ibu kota negara.

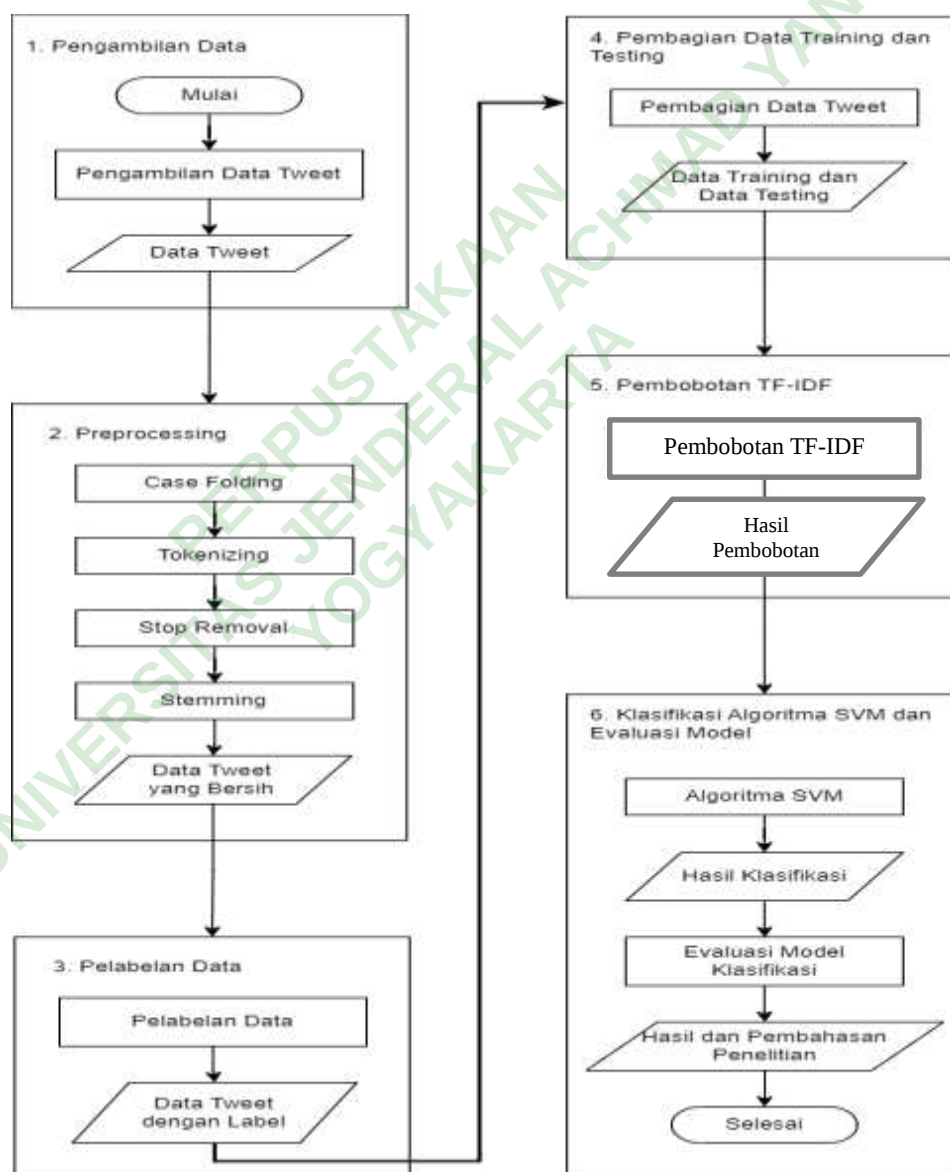
Alat penelitian yang digunakan adalah komputer dengan spesifikasi yang tepat untuk menjalankan sistem operasi, perangkat lunak untuk *entri* data, serta koneksitas internet. Sistem operasi aplikasi yang digunakan dalam pengembangan aplikasi ini adalah sebagai berikut:

1. Sistem Operasi: Windows 10
2. Jupyter Notebook

3. Python versi 3.11
4. Anaconda 32
5. Command Prompt 10.0

3.2 JALAN PENELITIAN

Penelitian ini memanfaatkan bahasa pemrograman Python, Comand Promt, dan Jupyter Notebook untuk melakukan pengambilan semua data yang akan disajikan. Jalannya penelitian dapat dilihat seperti pada **Gambar 3.1**



Gambar 3.1 Flowchart Jalannya Penelitian

3.2.1 Pengambilan Data

Tahap pengambilan data *tweet* yang berkaitan tentang perpindahan Ibu Kota Negara menggunakan *keyword* ibu kota negara yang diambil menggunakan teknik *snsrape* pada Jupyter Notebook selama periode 1 Juli hingga 2 Agustus 2023. Berikut kode – kode *library* pengambilan data *tweet* dapat dilihat di bawah ini:

```
pip install snsrape
```

library di atas merupakan kode yang digunakan untuk menginstal *snsrape* menggunakan *pip* yang nantinya akan digunakan untuk memanggil data *tweet*. Langkah selanjutnya setelah *snsrape* terinstall adalah melakukan pemanggilan *library*nya, berikut kode pemanggil *library* dengan cara mengimport *library* yang akan digunakan pada tahap selanjutnya, seperti yang terlihat di bawah ini:

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

Kode pengambilan data *tweet* pada Twitter dapat dilihat di bawah ini:

```
import os
import snsrape

# Using OS library to call CLI commands in Python
os.system("snsrape --jsonl --max-results 100000 --since 2022-12-01 twitter-search \"ibukotabaru until:2023-17-07\" > ibukn.json")

df= pd.read_json('ibukn.json', lines=True)
df.to_csv('ibukn.csv')

import pandas as pd
tweets_df = pd.read_csv('ibukn.csv')
df = tweets_df[["id", "username", "date", "content", "url", ]]
df
```

Kode tersebut digunakan untuk mengambil data *tweet* dan menyimpannya atau mengekspor kedalam bentuk file csv. Hasil data *tweet* yang telah diekspor dapat dilihat pada tabel 3.1 di bawah ini:

Tabel 3.1 Data Tweet

No.	Data Tweet
1	1666595850397827072,YogiIsk99022685,2023-06-07 23:59:49+00:00,@detikcom Promosiin aja bisa bikin negara di ikn,https://twitter.com/YogiIsk99022685/status/1666595850397827072
2	1666595819620032514,DJshyam1,2023-06-07 23:59:41+00:00,"@geloraco Kenapa yang diajak tinggal di ikn orang Singapura,bukan orang Indonesia,nanti itu kaya PIK,gak ada nama Darsono,Dadang,Rojali.",https://twitter.com/DJshyam1/status/166659581 9620032514
3	1666595457626435586,YogiIsk99022685,2023-06-07 23:58:15+00:00,@detikcom Wkwkw bisa bikin negara di ikn 😂😂😂,https://twitter.com/YogiIsk99022685/status/166659545762643558 6
4	1666592716569079809,D002832,2023-06-07 23:47:22+00:00,"@CNNIndonesia @jokowi Hahahaha, presiden paling kocak,ah... Wrong speech.... Ancur..ancurrrr... sogokan pasir laut ke Singapura, biar pengusaha Singapura mau investasi di IKN...ini lah kesalahan kita telah memilih pemimpin dari pedagang mebel,..😂😂😂",https://twitter.com/D002832/status/16665927165690798 09
5	1666592714069254144,NsarHaq,2023-06-07 23:47:21+00:00,"@msaid_didu Waktu tol kalian juga gini, koar2 melampiasikan kedengkian nyatanya lancar jaya dan mayoritas rakyat menikmati.
6	Saat ini ikn on progress tunggu saja akan jadi luar biasa.
7	Dan kalian akan tetap mengumpat sanasini",https://twitter.com/NsarHaq/status/1666592714069254144
8	1666591261812490244,nar_3300,2023-06-07 23:41:35+00:00,Daya pikat IKN apa ya ? 😂,https://twitter.com/nar_3300/status/1666591261812490244
9	yaahh bolak balik bginilah nasib wong cilik ni negri tercinta ini☹️",https://twitter.com/ikhwanjoeari/status/1666588921713815553

10	1666580252540411904,MellsDentist,2023-06-07 22:57:50+00:00,"@yazid_baidhawie @KakekHalal Pilih yg estafet menjalankan program2 pak Jokowi yg berguna bagi rakyat Indonesia krn program2 pak Jokowi terasa banget sampai ke bawah, bukan Java Centris saja. Pilih yg melanjutkan proyek IKN Nusantara. Jokowi di hati Ganjar dinanti. Bismillah semoga niat baik diijabah Allah😊 https://t.co/wRtRQI9AMs ", https://twitter.com/MellsDentist/status/1666580252540411904
----	--

Pada Tabel 3.1, data *tweet* yang berhasil diambil masih memiliki banyak bagian yang tidak relevan dalam proses klasifikasi sentimen, untuk mengubahnya menjadi relevan maka bagian yang tidak relevan perlu dihapus atau dibersihkan melalui proses *preprocessing*.

3.2.2 Pengelolaan Data (Preprocessing)

Tahap ini dilakukan pada data mentah lalu diolah menjadi data yang siap digunakan. Tahap *preprocessing* menggunakan langkah-langkah dan kode *library* yang ditampilkan untuk melakukan *preprocessing* data sebagai berikut:

```
import numpy as np
import pandas as pd
from pandas import DataFrame
df = pd.read_csv("ibukota.csv")
df = df[["id", "username", "date", "content", "url",]]
df

import nltk, re, string
from pandas import DataFrame
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

data = pd.read_csv
(r'ibukota.csv',nrows=10,names=['date','content']).dropna()
df.drop(['date'], axis=1, inplace=True)
```

Seperti kode di atas, fungsi *library numpy* yaitu mengolah data yang berbentuk numerik atau angka, *library pandas* adalah mengimpor dan mengelola data menjadi terstruktur, untuk menganalisis *text* menggunakan *library nltk*, sedangkan *library*

string digunakan untuk memproses data teks dalam Natural Language Processing (NLP) dalam kata lain yaitu bahasa alami, *library regex* atau *re* berfungsi untuk membantu mendefinisikan sebuah pola pencarian.

3.2.2.1 Text Cleaning

Text Cleaning atau membersihkan data adalah menyeleksi kembali data mentah dengan menghapus data yang tidak akurat dan tidak relevan. Kode untuk *text cleaning* sebagai berikut:

```
def
emoji_remove(ulasan):
if not ulasan:
return ''
t = ulasan.encode('ascii',
'ignore').decode('utf-8')
return t
```

Kode diatas digunakan untuk melakukan untuk menghapus *emoticon* yang terdapat pada dokumen.

```
def
url_remove(ulasan)
: if not ulasan:
return ''
t =
re.sub(r'http\S+', '',
ulasan)
return t
```

Kode diatas digunakan untuk menghapus *url* yang terdapat pada dokumen.

```
def hashtag_remove(ulasan):

if not ulasan: return
''
reg = r'#\S+'
t =
re.sub(reg, ' ',
ulasan)
return t
```

Kode diatas digunakan untuk menghapus *hashtag* yang terdapat pada dokumen.

```
def
remove_user(ulasan)
: if not ulasan:
```

```

return '' t =
re.sub(r'@[^\s]+', '',
ulasan) return t

```

Kode diatas digunakan untuk menghapus *mention* yang terdapat pada dokumen.

```

def
punc_remove(ulasan)
: if not ulasan:
return ''
t = re.sub(r'[\w\s]+', '', ulasan) return t

```

Kode di atas adalah kode untuk melakukan proses *text cleaning*, berikut adalah tabel hasil tahap *text cleaning* yang sudah bersih dari *emoji*, *hashtag*, tanda baca, *mention*, simbol, dan *url*, sebagai berikut seperti pada tabel 3.2 :

Tabel 3.2 Data Cleaning

No.	Data Tweet
1	Promosiin aja bisa bikin negara di ikn
2	Kenapa yang diajak tinggal di ikn orang Singapurabukan orang Indonesiananti itu kaya PIKgak ada nama DarsonoDadangRojali
3	Wkwkw bisa bikin negara di ikn
4	Hahahaha presiden paling kocakah Wrong speech Ancurancurrrr sogokan pasir laut ke Singapura biar pengusaha Singapura mau investasi di IKNNini lah kesalahan kita telah memilih pemimpin dari pedagang mebel
5	Waktu tol kalian juga gini koar2 melampiaskan kedengkian nyatanya lancar jaya dan mayoritas rakyat menikmati
6	Saat ini ikn on progress tunggu saja akan jadi luar biasa
7	Dan kalian akan tetap mengumpat sanasini
8	Daya pikat IKN apa ya
9	yaahh bolak balik bginilah nasib wong cilik ni negri tercinta ini
10	Pilih yg estafet menjalankan program2 pak Jokowi yg berguna bagi rakyat Indonesia krn program2 pak Jokowi terasa banget sampai ke bawah bukan Java Centris saja Pilih yg melanjutkan proyek IKN Nusantara Jokowi di hati Ganjar dinanti Bismillah semoga niat baik diijabah Allah

3.2.2.2 Case Folding

Fungsi tahap ini adalah mengubah huruf besar menjadi huruf kecil (*lowercase*) atau mengubah huruf kecil menjadi huruf besar (*uppercase*) agar mudah dalam mencocokkan kata. Berikut adalah kode program *case folding* seperti di bawah ini:

```
def lowercase():
    lower_word = df['content'].str.lower()
    return lower_word

lower_tweet = lowercase()

print(lower_tweet)
```

Hasil dari kode diatas semua data *tweet* diubah menjadi *lowercase* seperti tabel 3.3 di bawah ini:

Tabel 3.3 Data Tabel Lowercase

No.	Data Tweet
1	promosiin aja bisa bikin negara di ikn
2	kenapa yang diajak tinggal di ikn orang singapurabukan orang indonesiananti itu kaya pikgak ada nama darsonodadangrojali
3	wkwkw bisa bikin negara di ikn
4	hahahaha presiden paling kocakah wrong speech ancurancurrrr sogokan pasir laut ke singapura biar pengusaha singapura mau investasi di iknini lah kesalahan kita telah memilih pemimpin dari pedagang mebel
5	waktu tol kalian juga gini koar2 melampiasikan kedengkian nyatanya lancar jaya dan mayoritas rakyat menikmati
6	saat ini ikn on progress tunggu saja akan jadi luar biasa
7	dan kalian akan tetap mengumpat sanasini
8	daya pikat ikn apa ya
9	yaahh bolak balik bginilah nasib wong cilik ni negri tercinta ini
10	pilih yg estafet menjalankan program2 pak jokowi yg berguna bagi rakyat indonesia krn program2 pak jokowi terasa banget sampai ke

bawah bukan java centris saja pilih yg melanjutkan proyek ikn nusantara jokowi di hati ganjar dinanti bismillah semoga niat baik diijabah allah

3.2.2.3 Stopword Removal

Untuk mengurangi *level noise*, waktu pemrosesan dan ukuran *index* dapat dilakukan penghilangan *stopword*. Sedangkan *stopword* sendiri yaitu kata umum yang biasanya muncul akan tetapi tidak memiliki makna tertentu seperti kata penghubung dan, serta, setelah, yang, juga, dan lainnya. Tahap *stopword removal* digunakan untuk menghapus kata-kata yang tidak memiliki arti penting. Berikut merupakan kode instalasi *library sastrawi* serta kode proses *stop removal* seperti di bawah ini:

```
!pip install Sastrawi
```

```
From Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory

factory = StopWordRemoverFactory()
more_stopword = ['rt', 'anjir', 'anjing', 'monyet', 'jancok',
'tolol', 'bego', 'goblok', 'njing']
stopword = factory.create_stop_word_remover()
stopwords = factory.get_stop_words()+more_stopword
print(stopwords)
```

Kode tersebut digunakan untuk menyeleksi lalu menampilkan kata yang harus dihilangkan atau tidak penting, seperti pada tabel 3.4 berikut:

Tabel 3.4 Hasil Stopword Removal

Hasil library Sastrawi yang tidak memiliki makna
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehingga', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'sekitar', 'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'dalam', 'bi

```
sa', 'bahwa', 'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'anda', 'begitu', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'dimana', 'kemana', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai', 'sedangkan', 'selagi', 'sementara', 'tetapi', 'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nantinya', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimana', 'rt', 'anjir', 'anjing', 'monyet', 'jancok', 'tolol', 'bego', 'goblok', 'njing']
```

Apabila kode di atas digunakan untuk menampilkan kata yang dihilangkan dari data *tweet*, maka berikut untuk menghilangkan kata-kata dalam tabel di atas menggunakan kode seperti di bawah ini:

```
def removeStopWords(content):
    clean_word_list = [word for word in content.split() if word
not in stopwords]
    return clean_word_list

stopwords_tweet = lower_tweet.apply(removeStopWords)

print(stopwords_tweet)
```

3.2.2.4 Stemming

Pada stemming bentuk kata diubah menjadi kata dasar dalam arti lain yaitu mencari *root* dari tiap kata hasil dari tahap *filtering*. Pada tahap stemming masih menggunakan kode *library sastrawi*. Berikut adalah kode tahap stemming:

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return stemmer.stem(term)
```

```

term_dict = {}

for document in stopwords_tweet:
    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_tweet = stopwords_tweet.apply(get_stemmed_term)

print(stem_tweet)

```

3.2.3 Pelabelan Data

Pelabelan data yaitu mengkategorikan kumpulan data menjadi kelompok tertentu ketika data *tweet* yang sudah bersih. Dalam tahap ini kata atau kalimat akan diberi label sehingga dapat *dianalisis* bersifat positif atau negatif. Cara pelabelan manual dihitung dengan menggunakan *Textblob* untuk mendapatkan nilai *polarity* <0 yang dilabeli sebagai sentimen negatif, nilai *polarity* $=0$ dilabeli sentimen netral, dan nilai *polarity* >0 dilabeli sebagai sentimen positif.

Tabel 3.5 Tabel Pelabelan Manual

No.	Label	Content
-----	-------	---------

1.	Negatif	benar blinger ikn sdh murah meriah hgb th bebas pjg th diskon pjg tanah jd hak milik itupun blm cukup sekarang mengajak wn singapura tinggal luar biasa pak joko apa dilakukan ikn walaupun dg menggadai sma sdm negara
2.	Negatif	daya pikat ikn apa
3.	Negatif	duit kan gak punya warga negara rupiah bs direalkan yang pd naik haji dolar dirupiahkan bandara kertajati yang msh cari koneksitas pelabuhan patimban siap jaga berlabuhnya produk otomotif diributin gak sabaran ikn jngn hny menunggu pemaidtho
4.	Positif	jokowi investor singapura investasi indonesia aman di ikn
5.	Negatif	narasi perpanjangan kekuasaan rangka pembangunan ikn kelucuan yang didustakan
6.	Negatif	bukan jualan prestasi malah jualan ilmu sulap eh sihir coba dulu buat nyihir ikn biar jadi keliling cari investor
7.	Positif	welcome to ikn
8.	Negatif	hennn spertinya nnti pribumi kl masuk ikn hrs pake paspor n visa turis nihh yaahh bolak balik bginilah nasib wong cilik ni negri tercinta
9.	Positif	pemerintah indonesia ajak singapura dukung investasi ikn pembangunan menjanjikan
10.	Positif	mrk lbh menguntungkan impopasir laut indonesia utk memperluas negaranya teritorialnya negara mrk bertambah amp teritorial indonesia didepan wilayah yang mrk urug dg pasir laut impootomatis mengecil ditambah pemasukan pajak nantinya dibanding invest ikn

Tabel 3.5 di atas merupakan hasil dari pelabelan manual 1.000 dari 3.000 data tweet yang menghasilkan kelas lebih akurat lalu digunakan untuk tahapan training.

3.2.4 Training data

Pada tahap *training* data melakukan proses melatih data yang diawali dengan mengekstrasi data menggunakan TF-IDF dan selanjutnya melatih data untuk membentuk model klasifikasi analisis sentimen yang dapat digunakan secara otomatis. Dibawah ini merupakan kode program untuk mengimport algoritma SVM. Selanjutnya berikut kode program untuk memanggil data *training*:

```
def content_training(self, df):
    response = dict()
    X_train, X_test, y_train, y_test =
    train_test_split(df['normalization'].values,
    df['labelnew'].values,
```

```

test_size=0.2)

tf_vectorizer = TfidfVectorizer(min_df=4, max_df=0.3,
ngram_range=(1, 3))
    x_train_tfidf = tf_vectorizer.fit_transform(X_train)
    x_test_tfidf = tf_vectorizer.transform(X_test)

    vect = CountVectorizer()
    vect.fit(X_train)

    vect_path = "files/vectorizerkota-1.pkl"
    joblib.dump(vect, vect_path)

    data_train = vect.transform(X_train)
    data_test = vect.transform(X_test)

    model = SVC(kernel='linear', random_state=10)
    model.fit(data_train, y_train)

    prediction = model.predict(data_test)

    akurasi = accuracy_score(y_test, prediction)
    f1score = f1_score(y_test, prediction,
average='weighted') * 100
    precision = precision_score(y_test, prediction,
average='weighted') * 100
    recall = recall_score(y_test, prediction,
average='weighted') * 100

    response['x_train_tfidf'] = x_train_tfidf
    response['x_test_tfidf'] = x_test_tfidf
    response['akurasi'] = akurasi
    response['f1score'] = f1score
    response['precision'] = precision
    response['recall'] = recall
    response['train'] = len(X_train)
    response['test'] = len(X_test)

    return response

```

Kode di atas merupakan kode yang digunakan untuk proses *training* data untuk tahap klasifikasi SVM disimpan menggunakan format file berupa *pickle*.

3.2.5 Testing Data

Pada tahap *testing*, data yang tidak digunakan selama proses pelatihan digunakan untuk evaluasi seberapa baik model dapat melakukan prediksi pada data yang belum pernah terlihat atau data baru. Kode program yang digunakan untuk pemanggilan data *testing* seperti pada di bawah ini:

```
vect = CountVectorizer()
vect.fit(X_train)

vect_path = "files/vectorizerkota-1.pkl"
joblib.dump(vect, vect_path)

data_train = vect.transform(X_train)
data_test = vect.transform(X_test)

model = SVC(kernel='linear', random_state=10)
model.fit(data_train, y_train)

datas = df['normalization'].to_numpy()
text = np.array(datas, dtype=object)

sample = vect.transform(text)
prediction = model.predict(sample)

model_path = "files/svm_model.pkl"
joblib.dump(model, model_path)

df['model'] = prediction
positif = df.model.value_counts().positif
negatif = df.model.value_counts().negatif
total = positif + negatif
```

3.2.6 Klasifikasi

Pada tahap klasifikasi berfungsi untuk menampilkan hasil perhitungan menggunakan algoritma SVM menggunakan data bersih setelah tahap *preprocessing* sebanyak 986 data. Berikut kode program untuk menampilkan hasil klasifikasi sentimen:

```
def content_pengujian(self, df):  
    response = dict()  
    vect_path = "files/vectorizerkota-1.pkl"  
  
    datas = df['normalization'].to_numpy()  
    text = np.array(datas, dtype=object)  
  
    model_path = "files/svm_model.pkl"  
  
    test_vect = joblib.load(vect_path)  
    test_model = joblib.load(model_path)  
  
    sample = test_vect.transform(text)  
    prediction = test_model.predict(sample)  
  
    df['model'] = prediction  
    positif = df.model.value_counts().positif  
    negatif = df.model.value_counts().negatif  
    total = positif + negatif  
  
    response['df'] = df  
    response['total'] = total  
    response['positif'] = positif  
    response['negatif'] = negatif  
  
    return response
```

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA