

BAB 3

METODE PENELITIAN

Penelitian ini Metode menggunakan data Twitter untuk menganalisis sentimen negatif atau positif dengan memakai Pendekatan Algoritma SVM dengan membutuhkan pengumpulan data *tweet* yang terkait dengan *bullying* dan dilanjutkan dengan pengolahan data berupa *preprocessing* untuk mendapatkan hasil yang diperlukan. Informasi ini nantinya akan digunakan untuk mengumpulkan informasi sentimen tentang *bullying*.

3.1 BAHAN DAN ALAT PENELITIAN

Bahan-bahan penelitian yang mungkin dibutuhkan untuk menganalisis sentimen *bullying* negara pada media sosial Twitter antara lain:

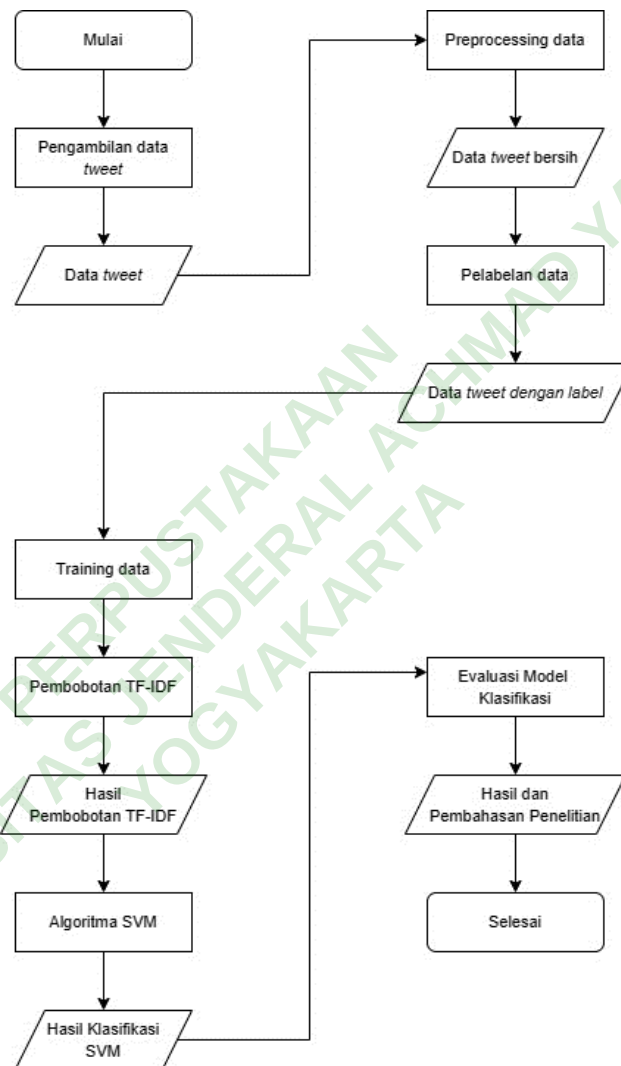
Dalam penelitian ini peneliti membutuhkan data *tweet* dari Twitter dengan menggunakan kata “*bullying*”. Data *tweet* diambil dari bulan Mei-Juni 2023 dengan jumlah sebanyak 13.000 *tweet*. Selain itu data yang digunakan hanya 966 data dari keseluruhan dikarenakan terdapat beberapa bahasa asing dan yang paling banyak terlihat yaitu bahasa inggris dan kemudian saya *trasnlate* ke bahasa indonesia. Serta saya juga memperoleh informasi yang berhubungan dan relevan dengan penelitian ini dari referensi seperti jurnal ilmiah, buku, dan lain sebagainya.

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 8 (64 bit)
2. Jupyter notebook
3. Bahasa pemrograman Python 3.6.2
4. Microsoft excel 2010
5. Anaconda 3 (64 bit)
6. Flask 64 bit

3.2 JALAN PENELITIAN

Penelitian ini memanfaatkan bahasa pemrograman Python, *Comand Prompt*, 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 *bullying* menggunakan keyword " *bullying* " yang diambil menggunakan teknik *snsrape* pada *Jupyter Notebook*. 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 \"bullying until:2023-17-07\" > ibukn.json")

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

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

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

Tabel 3.1 Data Tweet

No.	Data Tweet
1	With NATO exercises it's a two way street and if Putin didn't bully his neighbors there wouldn't be a need but he ChechnyaGeorgiaBelarusMoldova and talk how part of Finland bongs to Russia
2	From the news telling you all the bullying and other horrible things before they go off the deep end, that kind of thing only helps reinforce the myth that abused children will become abusers and because the anti version of that myth is much more widespread, it makes this so
3	Skinny people feeling what fat ppl have been through since the beginning of time after bullying ppl for so long brings me joy

4	Kinako bullies Aya
5	Edwards Retaliation for bullying
6	really to bully a student for speaking the truth what is wrong with you
7	bullying that dey play
8	spotify ai straight up bullied me it told me it would play some oldies from my days to come and play fatty wap and kendrick it slapped tho
9	So want to say in all probability there is a chance and your partner blocked you ei ajanku girl and therefore your partner is bluffing first is the people partner is also involved in the business that the guy you are talking to is pretty sure it isn't so why don't you stay back:
10	Bullying someone is not cool, Percy is beautiful and amazing and beautiful, stop with this craziness

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 *Preprocessing*

Tahap ini dilakukan pada data mentah lalu diolah menjadi data yang siap digunakan. Tahap *preprocessing* meliputi *translate*, *text cleaning*, *case folding*, *stopword removal*, *stemming*, pelabelan data, *training* data, *testing* data, dan klasifikasi serta 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("bullying.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'bullying
csv',nrows=10,names=['date','content']).dropna()
df.drop(['date'], axis=1, inplace=True)
```

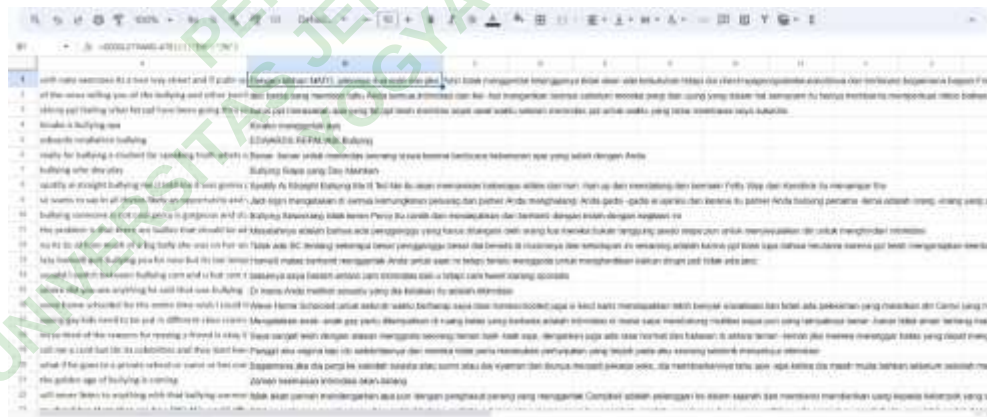
Seperti kode di atas, fungsi *library numpy* yaitu mengolah data yang berbentuk *numerik* atau angka, *library pandas* adalah mengimpor dan mengelola data menjadi terstruktur, untuk menganalisis *text* menggunakan *library nltk*, sedangkan *library string* digunakan untuk memproses data teks dalam Natural Language Processing (NLP) dalam kata lain yaitu bahasa alami, *library regex* atau *re* berfungsi untuk membantu mendefinisikan sebuah pola pencarian.

3.2.2.1 Translate

Translate atau terjemahan digunakan untuk penerjemah bahasa, cara yang digunakan untuk menerjemahkan file csv berbahasa inggris menjadi bahasa indonesia terdapat pada *excel online* dengan rumus sebagai berikut:

```
=GOOGLETRANSLATE(A3;"EN";"IN")
```

Dan hasil terjemahan dari menggunakan rumus di atas terdapat pada Gambar 3.2



Gambar 3.2 Hasil Terjemahan

3.2.2.2 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 url_remove(content):
    t = re.sub(r'http\S+', '', content)
    return t
def url (content):
    t = re.sub('((www\.[^\s]+)|(https?://[^\s]+))', 'URL', content)
    return t
def punc_remove(content):
    t = re.sub(r'^\w\s', '', content)
    return t
def rt_remove(content):
    t = re.sub(r'RT[\s]+', '', content)
    return t
def number_remove(content):
    t = re.sub('[0-9]+', '', content)
    return t
def slang_remove(content):
    t = re.sub(r'\\n', " ", content)
    return t
def regex_remove(content):
    t = re.sub("b'", " ", content)
    return t
def remove_user(content):
    t = re.sub('@[^\s]+', '', content)
    return t

```

```

def hashtag_remove(content):
    reg = "#(\w+:\w+/\w+)"
    return re.sub(reg, " ", content)

cleaned = []
def clean_text(content):
    for i in content:
        cleaned.append(url_remove(punc_remove(number_remove(remove_user(regex_remove(hashtag_remove(rt_remove(slang_remove(
            re.sub("[\n\r\t\xa0]", " ", i).strip()))))))))
    clean_text(df["content"])

```

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	Dengan latihan NATO itu adalah jalan dua arah dan jika Putin tidak menggertak tetangganya tidak akan ada kebutuhan tetapi dia ChechnyaGeorgiaBelarusMoldova dan berbicara bagaimana bagian dari bong Finlandia ke Rusia
2	Dari berita yang memberi tahu Anda semua intimidasi dan hal-hal mengerikan lainnya sebelum mereka pergi ke ujung yang dalam, hal semacam itu hanya membantu memperkuat mitos bahwa anak-anak yang dilecehkan akan menjadi pelaku kekerasan dan karena versi anti dari mitos itu jauh lebih luas, itu membuat ini begitu
3	Orang kurus merasakan apa yang telah dialami ppl gemuk sejak awal waktu setelah bullying ppl begitu lama membawa saya sukacita
4	Kinako menggertak Aya
5	Edwards Pembalasan bullying
6	benar-benar untuk menggertak seorang siswa karena berbicara kebenaran apa yang salah dengan Anda
7	bullying yang dey mainkan
8	spotify ai langsung menggertak saya itu mengatakan kepada saya itu akan memainkan beberapa oldies dari hari-hari saya yang akan datang dan bermain fatty wap dan kendrick itu menampar tho
9	Jadi ingin mengatakan dalam semua kemungkinan ada kesempatan dan patner Anda memblokir Anda ei ajanku gadis dan karena itu patner Anda menggertak pertama adalah orang-orang patner juga terlibat dalam bisnis yang pria Anda berbicara dengannya cukup yakin itu tidak jadi mengapa Anda tidak tinggal kembali
10	Menindas seseorang tidak keren, Percy cantik dan menakjubkan dan cantik, berhenti dengan kegilaan ini

3.2.2.3 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 Lowercase

No.	Data Tweet
1	dengan latihan nato itu adalah jalan dua arah dan jika putin tidak menggertak tetangganya tidak akan ada kebutuhan tetapi dia chechnyageorgiablarmoldova dan berbicara bagaimana bagian dari bong Ffnlandia ke rusia
2	dari berita yang memberi tahu anda semua intimidasi dan hal-hal mengerikan lainnya sebelum mereka pergi ke ujung yang dalam, hal semacam itu hanya membantu memperkuat mitos bahwa anak-anak yang dilecehkan akan menjadi pelaku kekerasan dan karena versi anti dari mitos itu jauh lebih luas, itu membuat ini begitu
3	orang kurus merasakan apa yang telah dialami ppl gemuk sejak awal waktu setelah bullying ppl begitu lama membawa saya sukacita
4	kinako menggertak aya
5	edwards pembalasan bullying
6	benar-benar untuk menggertak seorang siswa karena berbicara kebenaran apa yang salah dengan anda
7	bullying yang dey mainkan
8	spotify ai langsung menggertak saya itu mengatakan kepada saya itu akan memainkan beberapa oldies dari hari-hari saya yang akan datang dan bermain fatty wap dan kendrick itu menampar tho
9	jadi ingin mengatakan dalam semua kemungkinan ada kesempatan dan patner anda memblokir anda ei ajanku gadis dan karena itu patner anda menggertak pertama adalah orang-orang patner juga terlibat dalam bisnis yang pria anda berbicara dengannya cukup yakin itu tidak jadi mengapa anda tidak tinggal kembali
10	menindas seseorang tidak keren, percy cantik dan menakjubkan dan cantik, berhenti dengan kegilaan ini

3.2.2.4 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 Daftar 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', 'bisa', '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', 'nanti', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimana']

```
napun', '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)
```

Berikut tampilan hasil data stopword dari kode di atas ada pada Tabel 3.5

Tabel 3.5 Hasil Stopword Removal

No.	Data Tweet
1	['latihan', 'nato', 'jalan', 'arah', 'putin', 'menggertak', 'tetangganya', 'kebutuhan', 'chechnyageorgiabelarusmoldova', 'berbicara', 'bong', 'finlandia', 'rusia']
2	['berita', 'intimidasi', 'mengerikan', 'pergi', 'ujung', 'membantu', 'memperkuat', 'mitos', 'anak', 'anak', 'dilecehkan', 'pelaku', 'kekerasan', 'versi', 'anti', 'mitos', 'luas']
3	['orang', 'kurus', 'merasakan', 'dialami', 'people', 'gemuk', 'bullying', 'people', 'membawa', 'sukacita']
4	['kinako', 'menggertak', 'aya']
5	['edwards', 'pembalasan', 'bullying']
6	['menggertak', 'siswa', 'berbicara', 'kebenaran', 'salah']
7	['bullying', 'dey', 'mainkan']
8	['spotify', 'ai', 'langsung', 'menggertak', 'memainkan', 'oldies', 'bermain', 'fetty', 'wap', 'kendrick', 'menampar', 'though']
9	['kesempatan', 'patner', 'memblokir', 'ei', 'ajanku', 'gadis', 'patner', 'menggertak', 'orang', 'orang', 'patner', 'terlibat', 'bisnis', 'pria', 'berbicara', 'dengannya', 'tinggal']

10	['menindas', 'keren', 'percy', 'cantik', 'menakjubkan', 'cantik', 'berhenti', 'kegilaan']
----	---

3.2.2.5 Stemming

Pada stemming bentuk kata diubah menjadi kata dasar dalam arti lain yaitu mencari *root* dari tiap kata hasil dari tahap *filtering*. 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])
```

Berikut adalah tabel hasil tahap *stemming* yang sudah menjadi kata dasar dapat dilihat pada Tabel 3.6

Tabel 3.6 Hasil Stemming

No.	Data Tweet
1	['latih', 'nato', 'jalan', 'arah', 'putin', 'gertak', 'tetangga', 'butuh', 'chechnyageorgiabelarusmoldova', 'bicara', 'bong', 'finlandia', 'rusia']
2	['berita', 'intimidasi', 'keri', 'pergi', 'ujung', 'bantu', 'kuat', 'mitos', 'anak', 'anak', 'leceh', 'laku', 'keras', 'versi', 'anti', 'mitos', 'luas']
3	['orang', 'kurus', 'rasa', 'alami', 'people', 'gemuk', 'bullying', 'people', 'bawa', 'sukacita']

4	['kinako', 'gertak', 'aya']
5	['edwards', 'balas', 'bullying']
6	['gertak', 'siswa', 'bicara', 'benar', 'salah']
7	['bullying', 'dey', 'main']
8	['spotify', 'ai', 'langsung', 'gertak', 'main', 'oldies', 'main', 'fetty', 'wap', 'kendrick', 'tampar', 'though']
9	['sempat', 'patner', 'blokir', 'ei', 'ajanku', 'gadis', 'patner', 'gertak', 'orang', 'orang', 'patner', 'libat', 'bisnis', 'pria', 'bicara', 'dengan', 'tinggal']
10	['tindas', 'keren', 'percy', 'cantik', 'takjub', 'cantik', 'henti', 'gila']

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 3.7 Tabel Pelabelan Manual

No.	Label	Data Tweet
1.	Negatif	Dengan latihan NATO itu adalah jalan dua arah dan jika Putin tidak menggertak tetangganya tidak akan ada kebutuhan tetapi dia ChechnyaGeorgiaBelarusMoldova dan berbicara bagaimana bagian dari bong Finlandia ke Rusia
2.	Negatif	Dari berita yang memberi tahu Anda semua intimidasi dan hal-hal mengerikan lainnya sebelum mereka pergi ke ujung yang dalam, hal semacam itu hanya membantu memperkuat mitos bahwa anak-anak yang dilecehkan akan menjadi pelaku kekerasan dan karena versi anti dari mitos itu jauh lebih luas, itu membuat ini begitu
3.	Negatif	Orang kurus merasakan apa yang telah dialami ppl gemuk sejak awal waktu setelah bullying ppl begitu lama membawa saya sukacita

4.	Negatif	Kinako menggertak Aya
5.	Negatif	Edwards Pembalasan bullying
6.	Positif	benar-benar untuk menggertak seorang siswa karena berbicara kebenaran apa yang salah dengan Anda
7.	Negatif	bullying yang dey mainkan
8.	Negatif	spotify ai langsung menggertak saya itu mengatakan kepada saya itu akan memainkan beberapa oldies dari hari-hari saya yang akan datang dan bermain fatty wap dan kendrick itu menampar tho
9.	Positif	Jadi ingin mengatakan dalam semua kemungkinan ada kesempatan dan patner Anda memblokir Anda ei ajanku gadis dan karena itu patner Anda menggertak pertama adalah orang-orang patner juga terlibat dalam bisnis yang pria Anda berbicara dengannya cukup yakin itu tidak jadi mengapa Anda tidak tinggal kembali
10.	Positif	Menindas seseorang tidak keren, Percy cantik dan menakjubkan dan cantik, berhenti dengan kegilaan ini

Tabel 3.5 di atas merupakan hasil dari pelabelan manual 966 dari 13.000 data tweet yang menghasilkan kelas lebih akurat lalu digunakan untuk tahapan training.

3.2.4 Training data

Pada tahap training data di gunakan sebanyak 772 data dari 966 data keseluruhan dengan 303 data positif dan 469 data negatif untuk melakukan proses melatih data yang diawali dengan mekstrasi 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 digunakan selama proses pelatihan ada sebanyak 194 data dengan 82 data positif dan 112 data negatif 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:

```
def content_testing(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)

    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

response['df'] = df
response['total'] = total
response['positif'] = positif
response['negatif'] = negatif

return response

```

3.2.6 Klasifikasi

Pada tahap klasifikasi berfungsi untuk menampilkan hasil perhitungan menggunakan algoritma SVM menggunakan data bersih setelah tahap preprocessing sebanyak 966 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

