

BAB 3

METODE PENELITIAN

Pada Penelitian yang membahas tentang bansos ini merupakan penelitian analisis sentimen menggunakan metode *SVM* dengan menganalisa sentimen positif dan negatif yang diambil dari data Twitter. Pada Penelitian mengenai bansos ini dibutuhkan data postingan dari Twitter yang berkaitan tentang bansos. Pada tahapan berikutnya data yang sudah didapat kemudian diolah dengan tahapan *preprocessing* sehingga mendapatkan hasil data yang sesuai. Data yang sudah sesuai tersebut kemudian dapat digunakan untuk memetakan sentimen dari *warganet* di Twitter tentang bansos sehingga mendapatkan hasil informasi yang tepat mengenai proses pembagian bansos kepada masyarakat.

Penelitian ini diangkat dari latar belakang permasalahan yang ada, kemudian mengolah data yang sudah didapatkan sehingga dapat menentukan hasil sentimen yang tepat sesuai dengan informasi yang diperoleh. Berikut merupakan bahan, alat dan jalannya penelitian analisis sentimen tentang bansos serta tahapan penelitian guna menyelesaikan proses analisis sentimen menggunakan data *tweet*.

Beberapa metode untuk mendapatkan data atau informasi dalam pemecahan masalah. Berikut beberapa metode yang akan dilakukan antara lain:

1. Studi Pustaka

Studi pustaka atau metode pengumpulan data dengan menggunakan metode kepustakaan ini meliputi pengumpulan *literatur*, *paper*, jurnal, makalah maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi menulis khususnya analisis sentimen dengan menggunakan metode *SVM*.

2. Pengumpulan Data *tweet*.

Pada tahapan selanjutnya data yang diperoleh merupakan data yang didapat secara langsung di Twitter dengan menggunakan Jupyter Notebook terkait dengan pencarian #Bansos#bsu #blt #bansoscov19.

3. Perancangan.

Dalam tahapan ini ada beberapa hal yang akan dilakukan yaitu: desain alur pengambilan data, desain alur sistem dan metode, pemrograman.

4. Pengujian dan Evaluasi.

dalam tahapan ini dilakukan pengujian klasifikasi *SVM* untuk mengetahui letak kesalahan yang mungkin terjadi.

5. Penyusunan laporan

Dalam tahap ini penyusunan laporan dibuat sebagai dokumentasi yang berfungsi mempermudah untuk dipelajari dan di kembangkan oleh orang lain.

3.1 BAHAN PENELITIAN

Bahan penelitian akan menggali data dan informasi dari *tweet* yang ada di dalam Twitter terkait dengan bansos menggunakan *hashtag* #bansos, #bsu, #blt, #bansoscov19.

3.2 ALAT PENELITIAN

Pada Penelitian mengenai analisis sentimen membutuhkan alat yang digunakan untuk menjalankan sistem oprasi sehingga dibutuhkan komputer dengan spesifikasi cukup dan membutuhkan konektifitas internet. Aplikasi yang digunakan dalam analisis sentimen ini adalah:

1. Windows 8.1 s/d terbaru
2. Anaconda versi 3
3. Jupyter Notebook
4. Bahasa Pemrograman Python 3.7

Library yang digunakan dalam penelitian ini adalah sebagai berikut :

1. *Numpy* : Digunakan untuk membuat objek kelas *array* tunggal dan multidimensi
2. *Pandas* : Sebagai *library* yang mengatur tata letak data sehingga mudah dicari secara intuitif
3. *Nltk* : Digunakan untuk memudahkan pemrosesan teks seperti melakukan *classification*, *tokenization*, *stemming*, *tagging*, *parsing*, dan *semantic reasoning*.
4. *Sastrawi* : *Sastrawi* merupakan *library* sederhana yang dapat mengubah kata berimbuhan bahasa Indonesia menjadi bentuk dasarnya.
5. *Sklearn* : Digunakan untuk membantu melakukan processing data ataupun melakukan *training* data untuk kebutuhan machine learning atau data science.
6. *Pickle* : Digunakan untuk menyimpan dan membaca data ke dalam atau dari suatu file berformat .pkl.
7. *Snsrape* : Digunakan sebagai *library* pengambilan data Twitter tanpa menggunakan *API*.
8. *Seabron* : Dapat digunakan untuk menentukan hubungan antara dua variabel
9. *Matplotlib* : Untuk kepentingan visualisasi data yang dapat berjalan di *multi-platform* dengan menggunakan *Numpy*

3.3 JALAN PENELITIAN

Pada Penelitian mengenai bansos ini merupakan penelitian sentimen positif dan negatif pada Twitter. Pada penelitian ini metode yang digunakan adalah *SVM*. Pada tahapan awal mengambil data *tweet* yang diambil dari Twitter mengenai bansos , pada tahap selanjutnya adalah pengolahan data yaitu *preprocessing* data untuk mendapatkan hasil yang sesuai.

Data *preprocessing* merupakan teknik data *mining* yang melibatkan data mentah menjadi format data yang mudah dimengerti. Berikut ini langkah-langkah dalam metode penelitian, yaitu :

3.3.1 Pengambilan Data.

Pengambilan data merupakan tahapan diambilnya sebuah data yang berkaitan dengan bansos dari Twitter untuk mengambil data *tweet* mengenai bansos dengan memakai *software* Anaconda Prompt kemudian dijalankan menggunakan Jupyter Notebook yang selanjutnya hasilnya ditampilkan di dalam format MS.Excel. Pengambilan data *tweet* mengenai bansos yaitu dengan menggunakan hastag “#Bansos #bsu #blt #bansoscov19” selama periode 01 Januari 2020 – 22 Juli 2022 dengan mendapatkan jumlah 15.320 *tweet*. Dalam pengambilan data ini memakai *library* yang berada di Python. *Snsrape* merupakan *library* yang dipakai. *Library* pada Python ini digunakan karena tidak memerlukan *API* Twitter sehingga tidak perlu mengakses akun *Developer* Twitter. *Tweet* yang telah didapat dalam pengambilan data ini berformat *csv* yang kemudian akan dimanipulasi dengan menggunakan *library pandas* untuk pengambilan *text* dan *datetime*. Berikut merupakan kode untuk mengimport *library* pengambilan data yang dapat dilihat dalam Gambar 3.1.

```
: # Setting variables to be used below
maxTweets = 25000

# Creating list to append tweet data to
tweets_list2 = []

# Using TwitterSearchScraper to scrape data and append tweets to list
for i,tweet in enumerate(sntwitter.TwitterSearchScraper('bansos since:2020-01-01 until:2022-07-22 lang:id').get_items()):
    if i>maxTweets:
        break
    tweets_list2.append([tweet.date, tweet.id, tweet.content, tweet.username])
```

Gambar 3. 1 Library Pengambilan Data

Dalam pengambilan data menggunakan *library* yang akan disimpan ke dalam format *csv* memiliki beberapa bentuk variasi seperti *URL*, *hashtag*, *datetime*, *username* sehingga harus dilakukannya tahapan pengolahan data agar menghasilkan data *tweet* yang diinginkan. Berikut ini merupakan data *tweet* yang sudah diambil dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Tahapan Pengambilan Data Twitter

No.	Datetime	Text
1.	2022-06-28 07:12:38+00:00	Soal Dugaan Korupsi Bansos di Ngaringan Grobogan, Kapolres: Ada Laporan Serupa Selengkapnya klik tautan berikut ini. #grobogan #korupsibansos #bansos https://t.co/Hsqp3g74OF
2.	2022-06-25 00:46:07+00:00	Mending makan #bakso dari pada makan #bansos #StabilkanHargaPangan #StabilkanHargaPangan
3.	2022-06-24 01:36:30+00:00	Minggu Keempat Bulan Juni 2022, Bansos PKH dan BPNT Cair Besok? Login https://t.co/JKuOa7GA0f Adakah Namamu? https://t.co/Z4QwjShSD5 #bansos #pkh #bpnt #bsu2022
4.	2022-06-23 13:03:08+00:00	BLT Subsidi Gaji ternyata bisa gagal cair walau sudah memenuhi syarat, cek penyebabnya https://t.co/o7T8LP1siU #BLT #bantuan #bansos #SubsidiGaji #beritaterkini #beritaterbaru
5.	2022-06-23 09:38:25+00:00	Sudah tahu apa itu bansos PBI sebenarnya? simak penjelasan lengkap dan cara mendapatkannya https://t.co/t0kuBZ8oLy #bansos #bantuansosial #bantuan #PBI #beritaterkini #beritaterbaru
6.	2022-06-23 05:18:46+00:00	Update Daftar Penerima Bansos PKH Juni 2022, Cek di https://t.co/JKuOa7GA0f dan

		Dapatkan Rp3 Juta https://t.co/fx4ia6W78s #pkh #bansos #bpnt #bantuan #sekolahkusejahtera
7.	2022-06-23 00:23:38+00:00	Periode Akhir Bulan Juni 2022 Bansos PKH dan BPNT Cair? Login di https://t.co/JKuOa7GA0f Sekarang https://t.co/D9sC2OQGni #bansos #pkh #bpnt #bantuan #bsu2022
8.	2022-06-22 11:59:01+00:00	Bantuan Sosial Program Keluarga Harapan Tahap Kedua akan cair pada bulan Juni 2022 ini. Penyaluran Bansos PKH sudah memasuki tahap yang ke dua pada Juni 2022. #pkh #bansos https://t.co/kuYko36SSD
9.	2022-06-22 09:52:00+00:00	Hati-Hati! Ada Pesan Berantai Pendaftaran Penerima Bansos PKH Tahap 2 #kementriansosial #PKH #PesanBerantai #Bansos @KemensosRIKlikðŸ‘† https://t.co/aDrgUnGTBX
10.	2022-06-22 09:18:58+00:00	Penyelarasan Regulasi BSU 2022 Dikebut, Pencairan Segera Dilakukan #BSU #Bansos #subsidiupah #BSUCair #pencairan #kemnaker https://t.co/CCj8apzmLV

Pada Tabel 3.1 menunjukan hasil dari pengambilan pada data *tweet* masih banyaknya karakter yang tidak berfungsi untuk melakukan analisis sentimen. Beberapa *tweet* masih memiliki karakter yang tidak perlu untuk digunakan sehingga perlu melalui tahapan *preprocessing* data sehingga menghasilkan data yang lebih baik.

3.3.2 Data Preprocessing

Data *preprocessing* merupakan teknik data mining yang mentransformasikan dataset menjadi format data yang mudah untuk dimengerti. Dalam melakukan *preprocessing* dataset memerlukan *library* dari Python untuk menyelesaikan beberapa perintah. Pada *preprocessing* data sebelumnya perlu mengimport beberapa *library* yang bisa dilihat dalam Gambar 3.2

```
In [70]: import numpy as np
import pandas as pd
from pandas import DataFrame
import nltk, emoji, re, string
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

Gambar 3. 2 Library Preprocessing Data

1. Data cleansing.

Pada tahapan data *cleansing* dipakai sebagai membersihkan menghilangkan karakter yang tidak digunakan dengan contoh untuk membersihkan (@), membersihkan hashtag (#), membersihkan URL, membersihkan tanda baca, membersihkan angka dan membersihkan *emoticon*. Kode yang digunakan dalam melakukan data cleansing dapat dilihat dalam Gambar 3.3

```

: # menghilangkan karakter

def url_remove(text):
    t = re.sub(r'https://\S+', '', text)
    return t

def punc_remove(text):
    t = re.sub(r'^\w\s', '', text)
    return t

def number_remove(text):
    t = re.sub(r'\d+', '', text)
    return t

def regex_remove(text):
    reg = "b',_"
    return re.sub(reg, " ", text)

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

cleaned = []

```

Gambar 3. 3 Tahapan *Data Cleansing*

Fungsi dari pemberian kata *remove* yang digunakan untuk menghapus beberapa karakter yang tidak digunakan seperti *hashtag*, *URL*, angka, *underscore*, tanda baca, dan *emoticon*.

2. *Tokenize*

Tokenize adalah langkah-langkah untuk membagi kalimat *tweet* menjadi kata-kata yang menyusun kalimat *tweet* tersebut. Pada tahapan ini berfungsi sebagai cara mempermudah analisa kata-kata tersebut, Kode yang digunakan dalam melakukan data cleansing dapat dilihat dalam Gambar 3.4

```

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

```

Gambar 3. 4 Kode *Tokenize*

3. *Case Folding*

Case folding berfungsi sebagai pengubahan penggunaan huruf-huruf kapital (*uppercase*) menjadi huruf-huruf kecil (*lowercase*).code yang digunakan pada *case Folding* dapat dilihat dalam Gambar 3.5.


```
# membuat semua huruf menjadi huruf kecil

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

lower_tweet = lowercase()

print(lower_tweet)
```

Gambar 3. 5 Tahapan Data *Cleansing*

Pada tahap *uppercase* diatas berfungsi untuk mengubah sebuah kata yang memiliki huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*).

4. *Stopword*

Pada tahapan *stopword* merupakan tahapan yang digunakan untuk menghapus sebuah kata-kata yang tidak digunakan atau tidak bermakna dan memiliki informasi data rendah dalam sebuah teks (“dan”, ”jikalau”, “pun”, “di”, “yang” dll). Kode yang dipakai untuk tahapan *stopword* dapat dilihat dalam Gambar 3.6

```
: # menghilangkan kata-kata yang tidak dibutuhkan

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
more_stopword = ['rt', 'bri', 'anjing', 'anjim', 'pantek', 'asu', 'jancok', 'kntl', 'kontol', 'bajingan', 'mmk', 'matam
stopword = factory.create_stop_word_remover()
stopwords = factory.get_stop_words()+more_stopword
print(stopwords)
```

Gambar 3. 6 Tahapan *Stopword*

5. *Stemming*

Stemming merupakan tahapan dari proses pencarian kata dasar dari suatu kata. Kata yang memiliki imbuhan seperti ber-, mem-, men-, -an,, per-, di-, ke, -kan, dan lain-lain akan dihapus. Kode yang dipakai untuk melakukan tahapan proses *stemming* dapat dilihat dalam Gambar 3.7

```
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)
```

Gambar 3. 7 Tahapan *Stemming*

6. *Normalization*

Normalization merupakan proses penyamaan pada kata-kata yang mempunyai kesalahan dalam penulisan atau menggunakan kalimat yang kurang lazim. Kode yang dipakai pada tahap *normalization* dapat dilihat dalam Gambar 3.8

```

normalizad_word = pd.read_excel("normalization.xlsx")
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_tweet = stem_tweet.apply(normalized_term).str.join(" ")
print(normal_tweet)

```

Gambar 3. 8 kode tahapan *normalization*

Pada tahap *normalization* dilakukannya perubahan kata-kata yang belum sesuai dari daftar kata yang sudah dibuat di dalam file Excel dengan nama *normalization.xlsx* kemudian kata tersebut diubah menjadi *dataframe* dengan nama *normal_tweet* kemudian mendapatkan hasil *tweet* yang sudah melalui tahap *preprocessing*. Contoh data *tweet* yang ada dalam *normalization.xlsx* yang digunakan untuk *dataset* sebagai penyamaan kata-kata yang kurang sesuai dapat dilihat dalam Tabel 3.2.

Tabel 3. 2 Data *Normalization*

Sebelum	Sesudah
ttg	tentang
dr	dari
gua	saya
seneng	senang
pake	pakai
bgt	banget
yg	yang
utk	untuk
liat	lihat

Pada Tabel 3.2 merupakan hasil dari data *tweet* yang sudah melewati tahapan *preprocessing*, menjadikan kata lebih terstruktur dan sempurna sehingga dapat dipakai dalam perhitungan ditahapan selanjutnya.

3.3.3 Pelabelan Manual

Tahapan pelabelan secara manual merupakan tahapan pemberian label positif atau negatif kepada kata secara manual di dalam suatu dokumen sehingga data *tweet* tersebut bisa dilakukan analisis selanjutnya mengenai sifatnya negatif atau positif. Dari *tweet* yang telah melalui tahapan *preprocessing* dan dilakukan tahapan pelabelan secara manual maka diperoleh sebesar 1.427 data *tweet* dengan jumlah 714 data *tweet* negatif dan 713 data *tweet* positif. Hasil dari tahap pelabelan secara manual dapat dilihat dalam Tabel 3.3

Tabel 3. 3 Contoh *tweet* Pelabelan Manual

Text	Score	Sentimen
bupati grobogan sumarnigrobogan resmi buka bimbing teknis damping keluarga harap dinas sosial kab grobogan hotel grand master rabu jul moh sumarsono dinsosgrob	1	positif
cegah stunting beri stimulasi cukup tepat mohon bantu like subscribe nya saudarasaudara	1	positif
kamis bhabin jagal muh syarifudin rakor damping pkh kapanewon banguntapan kaur kesra kalurahan rakor salur pkh tepat hak terima	1	positif
sangat inspirasi muncul mental fighter suka rintis konsisten jalan usaha walaupun kondisi sulit suka semua libat penuh angkat kemenbumn	1	positif

kemensosri bnpb indonesia selamat siang pak jokowi ibu terima pkh bpnt nama tera online terima bantu bulan desembersekarang henti alas jelas camat kk henti paks	1	positif
apa kata soal bansos nuraeni terima manfaat pm program keluarga harap pkh asal kalong jawa tengah cerita bagaimana guna dana bantu pkh nya cukup butuh seharihari simak video lengkap	-1	negatif
pemkab tangerang salur bansos masyarakat warga isoman salam sehat semua jangan pernah putus doa pandemi segera akhir aamiin	-1	negatif
salur program keluarga harap pkh tahap keluarga terima manfaat kpm provinsi indonesia laku tahap sejak awal juli	-1	negatif
bansos pkh beri per tiga bulan sesuai kriteria tentu	-1	negatif
pasang stiker warga miskin peberima tidak perlu laku kalau data laku baik	-1	negatif

Dari Tabel 3.3 dapat dilihat data *tweet* dengan sentimen positif diberi label dengan nilai 1 sedangkan *tweet* dengan sentimen negatif diberi label menggunakan nilai -1. Pada tahap pelabelan ini bertujuan untuk mengetahui sentimen positif dan negatif kemudian akan dilakukan tahap penghitungan akurasi.

3.3.4 Training Data

Tahapan *training* data merupakan tahapan *training* dengan memakai metode *SVM*. Pada *training* data dimulai dengan tahapan ekstraksi kata pada suatu data dengan memakai *TF-IDF*, selanjutnya akan dilakukan tahap pemodelan klasifikasi yang bertujuan untuk menentukan sentimen secara otomatis. Contoh kata yang akan dilakukan tahap penghitungan *TF-IDF* secara manual dapat dilihat pada Tabel 3.4.

Tabel 3. 4 Dokumen *TF-IDF*

Dokumen	Text
d1	minggu tiga bulan juli benar pkh tahap cair besok
d2	pkh tahap cair hari begini kategori terima pkh
d3	bansos pkh bpnt cair juli cek syarat salur
d4	batas tanggal cair bpnt juli cek link
d5	syarat terima pkh bpnt juli segera cair

Tabel 3.4 adalah contoh kata yang akan dilakukan tahap pembobotan *TF-IDF* secara manual dengan menggunakan dokumen dengan jumlah 5 yaitu dokumen 1, dokumen 2, dokumen 3, dokumen 4 dan dokumen 5. Dalam tahapan pembobotan *TF-IDF* secara manual ini akan menggunakan komponen seperti *T* yaitu *term* selanjutnya *d* yang berarti dokumen dokumen 1, dokumen 2, dokumen 3, dokumen 4 dan dokumen 5 serta *df* merupakan jumlah *term* yang sering muncul pada suatu dokumen. Contoh tahap penghitungan *TF* secara manual dapat dilihat dalam Tabel 3.5

Tabel 3. 5 Penghitungan *TF* Secara Manual.

Text	d1	d2	d3	d4	d5	Df
minggu	1					1
tiga	1					1
bulan	1					1
juli	1		1	1	1	4

benar	1					1
pkh	1	2	1		1	4
tahap	1	1				2
cair	1	1	1	1	1	5
besok	1					1
hari		1				1
begini		1				1
kategori		1				1
terima		1			1	2
bansos			1			1
bpnt			1	1	1	3
cek			1	1		2
link				1		1
syarat			1		1	2
segera					1	1
salur			1			1
batas				1		1
tanggal				1		1

Pada Tabel 3.5 menjelaskan tentang distribusi dari kemunculan *term* disetiap dokumen. *IDF* merupakan beberapa komponen *term* yang muncul atau d serta df merupakan jumlah dari kemunculan *term* dari semua dokumen. Contoh penghitungan *IDF* dapat dilihat dalam Tabel 3.6.

Tabel 3. 6 Penghitungan *IDF*

<i>Text</i>	<i>Df</i>		<i>d/df</i>	<i>idf</i>
minggu	1		5	0.69897
tiga	1		5	0.69897
bulan	1		5	0.69897
juli	4		1.25	0.09691
benar	1		5	0.69897
pkh	4		1.25	0.09691
tahap	2		2.5	0.39794
cair	5		1	0
besok	1		5	0.69897
hari	1		5	0.69897
begini	1		5	0.69897
kategori	1		5	0.69897
terima	2		2.5	0.39794
bansos	1		5	0.69897
bpnt	3		1.666667	0.221849
cek	2		2.5	0.39794
link	1		5	0.69897
syarat	2		2.5	0.39794
segera	1		5	0.69897
salur	1		5	0.69897
batas	1		5	0.69897
tanggal	1		5	0.69897

Pada Tabel 3.6 merupakan penjelasan mengenai hasil perhitungan dari *idf* secara manual dengan memakai rumus pada persamaan (1). Kemudian penghitungan *TF-IDF* dapat dilihat dalam Tabel 3.7.

Tabel 3. 7 Penghitungan *TF-IDF*

Text	d1	d2	d3	d4	d5
minggu	0.69897	0	0	0	0
tiga	0.69897	0	0	0	0
bulan	0.69897	0	0	0	0

juli	0.09691	0	0.09691	0.09691	0.09691
benar	0.69897	0	0	0	0
pkh	0.09691	0.19382	0.09691	0	0.09691
tahap	0.39794	0.39794	0	0	0
cair	0	0	0	0	0
besok	0.69897	0	0	0	0
hari	0	0.69897	0	0	0
begini	0	0.69897	0	0	0
kategori	0	0.69897	0	0	0
terima	0	0.39794	0	0	0.39794
bansos	0	0	0.69897	0	0
bpnt	0	0	0.221849	0.221849	0.221849
cek	0	0	0.39794	0.39794	0
link	0	0	0	0.69897	0
syarat	0	0	0.39794	0	0.39794
segera	0	0	0	0	0.69897
salur	0	0	0.69897	0	0
batas	0	0	0	0.69897	0
tanggal	0	0	0	0.69897	0

Pada Tabel 3.7 merupakan hasil dari penghitungan *TF-IDF* yang dilakukan dengan manual menggunakan perkalian dari *TF* dan *IDF*.

Pada penelitian ini dilakukannya tahap klasifikasi dengan memakai fitur ekstraksi *TF-IDF* yang akan memiliki hasil perhitungan secara otomatis memakai pembobotan *term* dalam dokumen data *training*. Dalam menghitung *TF-IDF* ini akan memakai *library* Python yaitu *Sklearn* dan *TfidfVectorizer* serta didukung dengan *library SVM* yang akan mendapatkan hasil secara otomatis. Hasil perhitungan *TF-IDF* secara otomatis dapat dilihat pada Gambar 3.9 dan Gambar 3.10.

(0, 294)	0.37210057499285104
(0, 366)	0.3185088726875455
(0, 845)	0.18160899972686018
(0, 208)	0.31207330465760563
(0, 277)	0.2632358323046435
(0, 413)	0.21837813132369027
(0, 184)	0.274200313777422
(0, 152)	0.31207330465760563
(0, 740)	0.35638186356772916
(0, 260)	0.3342275841126674
(0, 159)	0.3185088726875455
(1, 778)	0.38484241631146887
(1, 610)	0.27217435786236743
(1, 854)	0.38484241631146887
(1, 500)	0.38484241631146887
(1, 88)	0.15767352432946574
(1, 578)	0.3152986540336124
(1, 901)	0.2603342560644136
(1, 178)	0.3152986540336124
(1, 125)	0.2447166952641613
(1, 169)	0.3609189024946521
(2, 903)	0.2521443108233229
(2, 273)	0.29595774217679394
(2, 760)	0.31410851330175726
(2, 674)	0.3162481362602555

Gambar 3. 9 Perhitungan *TF-IDF* Awal

(998, 47)	0.2686018015221281
(998, 451)	0.2930763876449837
(998, 335)	0.27929301553002467
(998, 852)	0.28569074778315623
(998, 120)	0.3018117455366052
(998, 830)	0.3018117455366052
(998, 460)	0.2085796457463054
(998, 879)	0.2560314517050429
(998, 992)	0.2598664436305066
(998, 604)	0.24317843545427667
(998, 456)	0.19054191954028138
(998, 552)	0.18196339262842345
(998, 85)	0.1099876554708973
(998, 321)	0.2085796457463054
(998, 88)	0.1280353747568486
(998, 845)	0.15924870397837923
(999, 233)	0.3915511650959058
(999, 602)	0.40542125916278454
(999, 642)	0.3484667177079521
(999, 347)	0.36233681177483085
(999, 371)	0.40542125916278454
(999, 376)	0.3484667177079521
(999, 456)	0.2471968427304259
(999, 623)	0.24968348306128463
(999, 85)	0.1426909167144182

Gambar 3. 10 Perhitungan *TF-IDF* Akhir

Gambar 3.9 dan 3.10 menunjukkan hasil perhitungan *TF-IDF* secara otomatis dengan memanfaatkan *library Sklearn* dan *TfidfVectorizer*.

Kemudian setelah dilakukan proses perhitungan *TF-IDF* secara otomatis selanjutnya dilakukan penyesuaian model *dataset* dengan menggunakan algoritma *SVM*. Dalam membuat model ini digunakan sebuah fungsi untuk mempermudah pemanggilan dan untuk mengeksekusi pada tahapan selanjutnya sehingga lebih efisien dan efektif. Kode untuk membuat model bisa dilihat dalam Gambar 3.11.

```
SVM = SVC(C=1.0, kernel='linear', degree=3, gamma='auto')
text_pickle=SVM.fit(Train_X_Tfidf,Train_Y)
```

Gambar 3. 11 Pembuatan Model Klasifikasi

Setelah dilakukan proses pemodelan yang diberi nama *text_pickle* tahap selanjutnya kemudian disimpan kedalam file bernama *pickle* sehingga dapat dibuka kembali dan digunakan. Kode yang dipakai untuk menyimpan dan membuka file bernama *pickle* bisa dilihat dalam Gambar 3.12 dan dalam Gambar 3.13.

```
files = open('SVM_classifier_24.pickle', 'wb')
pickle.dump(text_pickle, files)
files.close()
```

Gambar 3. 12 Menyimpan File *Pickle*

```
model = open('SVM_classifier_24.pickle', 'rb')
svm_classifier = pickle.load(model)
svm_classifier
```

Gambar 3. 13 Membuka File *Pickle*

File *pickle* ini akan dipakai untuk eksekusi dalam data *Training* dengan jumlah 1000 *tweet* yang diambil sebagai *sample* dengan menggunakan *library smote* yang nantinya akan dilakukan prediksi Y baru menggunakan uji X. Kode yang digunakan untuk melakukan prediksi Y baru menggunakan uji X dapat dilihat pada Gambar 3.14.

```
SVM.fit(X_train_res, y_train_res)
y_pred = SVM.predict(X_test_vect)
```

Gambar 3. 14 Prediksi Y Baru

Fungsi *y_pred* ini digunakan untuk menghitung *Accuracy*, *F1 Score* dan *Confusion Matrix*. Kode untuk menghitung *Accuracy*, *F1 Score* dan *Confusion Matrix* dapat dilihat pada Gambar 3.15.

```
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))
```

Gambar 3. 15 Perhitungan *Accuracy*, *F1 Score* dan *Confusion Matrix*

Dengan Perhitungan *Accuracy*, *F1 Score* dan *Confusion Matrix* menghasilkan *training* data yang hasilnya dapat dilihat pada Gambar 3.16.

```
Average accuracy across folds: 84.40%
Average F1 score across folds: 84.37%
Average Confusion Matrix across folds:
[[90.9 12.9]
 [18.3 77.9]]
```

Gambar 3. 16 Hasil *Accuracy*, *F1 Score* dan *Confusion Matrix*.

Cross Validation merupakan suatu proses atau tahapan perhitungan akurasi berkali-kali (*k-fold*) dengan memakai parameter sama. Pada tahap ini sering dipakai dalam pencarian nilai akurasi dengan menggunakan beberapa kali percobaan sehingga akan mendapatkan tingkat performa data yang digunakan. Kode yang dipakai dalam melakukan *cross validation* bisa dilihat dalam Gambar 3.17.

```
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))
```

Gambar 3. 17 Perhitungan *Cross Validation*

Accs dan *f1s* merupakan hasil dari perhitungan *accuracy* dan *f1s* yang telah dilakukan pada Gambar 3.17.

3.3.6 Testing

Testing merupakan tahapan yang digunakan untuk mengukur keakuratan dari model yang telah dibangun, pada tahapan *testing* dipakai sebagai prediksi label dari data uji yang sudah ada. Pada tahapan *testing* ini akan dilakukan proses pelabelan secara manual dengan jumlah 400 data *tweet* dengan rincian sentimen negatif sebanyak 201 *tweet* dan positif 199 *tweet* nantinya pelabelan manual ini akan dibandingkan dengan pelabelan secara otomatis yang telah dilakukan oleh model *SVM*. Contoh pelabelan secara manual yang akan digunakan untuk perbandingan bisa dilihat pada Tabel 3.8.

Tabel 3. 8 Pelabelan Secara Manual

Text	Score	Sentimen
kube skema usaha kelompok yang diri atas terima manfaat pkh untuk kelola usaha kolektif usaha individual yang kelola lalu kelompok satu kelompok terima hibah juta untuk guna sebagai modal awal untuk kembang usaha	1	positif
responden yang dapat modul pk kata program tersebut manfaat dalam dukung giat ekonomi dalam kuat jejaring sosial dorong biasa tabung kelola keluar rumah tangga motivasi untuk mulai usaha	1	positif
intervensi mata cahari utama beberapa program lengkap rancang khusus keluarga untuk tingkat kapasitas produktif semua rancang kemensosri	1	positif
kualitas manfaat latih sangat variasi beberapa kursus latih tawar relevan konteks mata cahari tempat	1	positif
info tentang intervensi amp syarat untuk partisipasi bagi besar sampai dari mulut mulut amp lalu poster yang	1	positif

pajang kantor desa bagi besar keluarga milik jaring sosial amp hubung dengan kantor desa yang batas sehingga tidak terima informasi		
meski perintah lokal nasional tawar jumlah besar skema latih juru intervensi untuk atas hambat mata cari responden pernah terima dukung lalu intervensi tersebut	1	positif
sambut allaster cox deputy head of mission of australian embassy ucap apresiasi smeru mitra teliti australia hasil teliti kualitas penuh wawas	1	positif
presiden ri jokowi salur bantu sosial program keluarga harap pkh tahap i lapang rajawali cimahi jawa barat simak terus info lanjut kemensosri juliari batubara	1	positif
masuk kriteria cek informasi baru bantu pkh baru	-1	negatif
minggu tiga bulan juli benar pkh tahap cair besok	-1	negatif

Pada Tabel 3.8 menunjukkan *compound_score* -1 bersifat negatif sedangkan *compound_score* 1 bersifat positif kemudian dari data tersebut selanjutnya akan dibandingkan dengan data *tweet* yang telah diberi label secara otomatis dengan menggunakan model *SVM*. Kode untuk pelabelan secara otomatis dapat dilihat pada Gambar 3.18.

```
from sklearn import svm
classifier_linear = svm.SVC(kernel='linear')
classifier_linear.fit(train_vectors, y)
prediction_linear = classifier_linear.predict(test_vectors)
```

Gambar 3. 18 Pelabelan Secara Otomatis

Pada Gambar 3.18 *classifier_linear* merupakan pembuatan model untuk melakukan tahap *testing* data *tweet*. Contoh dari data *tweet* yang telah diberi label secara *otomatis* dengan menggunakan model *SVM* dapat dilihat pada Tabel 3.9.

Tabel 3. 9 Data Pelabelan Otomatis

Text	Score	Sentimen
kube skema usaha kelompok yang diri atas terima manfaat pkh untuk kelola usaha kolektif usaha individual yang kelola lalu kelompok satu kelompok terima hibah juta untuk guna sebagai modal awal untuk kembang usaha	1	positif
responden yang dapat modul pk kata program tersebut manfaat dalam dukung giat ekonomi dalam kuat jejaring sosial dorong biasa tabung kelola keluar rumah tangga motivasi untuk mulai usaha	1	positif
intervensi mata cahari utama beberapa program lengkap rancang khusus keluarga untuk tingkat kapasitas produktif semua rancang kemensosri	1	positif
kualitas manfaat latih sangat variasi beberapa kursus latih tawar relevan konteks mata cahari tempat	1	positif
info tentang intervensi amp syarat untuk partisipasi bagi besar sampai dari mulut mulut amp lalu poster yang pajang kantor desa bagi besar keluarga milik jaring sosial amp hubung dengan kantor desa yang batas sehingga tidak terima informasi	1	positif
meski perintah lokal nasional tawar jumlah besar skema latih juru intervensi untuk atas hambat mata cari responden pernah terima dukung lalu intervensi tersebut	1	positif
sambut allaster cox deputy head of mission of australian embassy ucap apresiasi smeru mitra teliti australia hasil teliti kualitas penuh wawas	1	positif
presiden ri jokowi salur bantu sosial program keluarga harap pkh tahap i lapang rajawali cimahi jawa barat simak terus info lanjut kemensosri juliari batubara	1	positif

masuk kriteria cek informasi baru bantu pkh baru	-1	negatif
minggu tiga bulan juli benar pkh tahap cair besok	-1	negatif

Pada Tabel 3.9 merupakan hasil dari pelabelan secara *otomatis* dengan menggunakan model *SVM*. Pada tahap selanjutnya data *tweet* yang diberi label secara manual dan data *tweet* yang telah diberi label secara otomatis digabungkan untuk melakukan pengecekan akurasi model *SVM*. Data *tweet* yang telah digabungkan bisa dilihat dalam Tabel 3.10.

Tabel 3. 10 Data Gabungan

Text	Actual	Predik
kube skema usaha kelompok yang diri atas terima manfaat pkh untuk kelola usaha kolektif usaha individual yang kelola lalu kelompok satu kelompok terima hibah juta untuk guna sebagai modal awal untuk kembang usaha	positif	positif
responden yang dapat modul pk kata program tersebut manfaat dalam dukung giat ekonomi dalam kuat jejaring sosial dorong biasa tabung kelola keluar rumah tangga motivasi untuk mulai usaha	positif	positif
intervensi mata cahari utama beberapa program lengkap rancang khusus keluarga untuk tingkat kapasitas produktif semua rancang kemensosri	positif	positif
kualitas manfaat latih sangat variasi beberapa kursus latih tawar relevan konteks mata cahari tempat	positif	positif
info tentang intervensi amp syarat untuk partisipasi bagi besar sampai dari mulut mulut amp lalu poster yang pajang kantor desa bagi besar keluarga milik jejaring sosial amp hubung dengan kantor desa yang batas sehingga tidak terima informasi	positif	positif

meski pemerintah lokal nasional tawar jumlah besar skema latih juru intervensi untuk atas hambat mata cari responden pernah terima dukung lalu intervensi tersebut	positif	positif
sambut allaster cox deputy head of mission of australian embassy ucap apresiasi smeru mitra teliti australia hasil teliti kualitas penuh wawas	positif	positif
presiden ri jokowi salur bantu sosial program keluarga harap pkh tahap i lapang rajawali cimahi jawa barat simak terus info lanjut kemensosri juliaribatubara	positif	positif
masuk kriteria cek informasi baru bantu pkh baru	negatif	negatif
minggu tiga bulan juli benar pkh tahap cair besok	negatif	negatif

Pada Tabel 3.10 menunjukkan pelabelan secara manual berada pada kolom *actual* sedangkan pelabelan secara otomatis berada pada kolom *predik*. Terdapat beberapa perbedaan yang ada dalam pelabelan secara manual dan otomatis. Selanjutnya akan ditentukan keterangan *TN*, *FN*, *TP* dan *TN*. Kode yang digunakan untuk menentukan *TN*, *FN*, *TP* dan *TN* bisa dilihat dalam Gambar 3.19.

```
df.loc[(df['Sentimen'] == 'positif') & (df['class'] == 'positif'), 'keterangan'] = 'TP'
df.loc[(df['Sentimen'] == 'positif') & (df['class'] == 'negatif'), 'keterangan'] = 'FP'
df.loc[(df['Sentimen'] == 'negatif') & (df['class'] == 'negatif'), 'keterangan'] = 'TN'
df.loc[(df['Sentimen'] == 'negatif') & (df['class'] == 'positif'), 'keterangan'] = 'FN'
```

Gambar 3. 19 Menentukan TN, FN, TP dan TN.

Pada tahapan ini akan dilakukan pemberian keterangan jika *actual* bersifat *negatif* dan *predik* bersifat *negatif* maka menjadi *True Negatif (TN)*, jika *actual* bersifat *positif* dan *predik* bersifat *negatif* maka *False Positif (FP)*, jika *actual* bersifat *positif* dan *predik* bersifat *positif* maka menjadi *True Positif (TP)* dan jika *actual* bersifat *negatif* dan *predik* bersifat *positif* maka menjadi *False Negatif (FN)*. Hasil pemberian keterangan dapat dilihat pada Gambar 3.20.

	Text	Sentimen	class	keterangan
0	bupati grobogan sumarnigrobogan resmi buka bim...	positif	positif	TP
1	cara daftar kartu prakerja gelombang lulus sma...	positif	positif	TP
2	seknas fitra sama spri koalisi reformasi lindu...	positif	positif	TP
3	cepat tangan miskin kemensos ri lalu damping p...	positif	positif	TP
4	bsu blt subsidi gaji bpjs ketenagakerjaan cair	positif	positif	TP
...
395	program tas miskin tahap rumah keluarga terima...	positif	negatif	FP
396	siapa yang menang tetep lanjut lengkap tambah ...	positif	positif	TP
397	kemensosri haha jauh nengok langit bantu pkh d...	positif	positif	TP
398	damping potong dana pkh beri sanksi tegas shar...	positif	positif	TP
399	salur bantu pkh pasiranjaya duga potong share ...	positif	positif	TP

Gambar 3. 20 Pemberian Keterangan

Gambar 3.20 menunjukkan hasil pemberian keterangan terhadap *sentimen actual* dan *sentimen predik*. Hasil pemberian keterangan akan dilakukan perhitungan *accuracy* yang dilakukan secara manual dengan menggunakan rumus pada persamaan (6). Hasil dari perhitungan *accuracy* dapat dilihat pada Gambar 3.21.

$$\text{Akurasi} = \frac{389}{400} = 0.9725$$

Gambar 3. 21 Hasil Perhitungan *Accuracy*

3.3.7 Hasil Klasifikasi

Setelah dilakukan tahap kalsifikasi pemodelan dengan memakai metode *SVM* yang telah diuji dengan *training* dan *testing* yang mendapatkan akurasi sesuai. Tahapan selanjutnya akan dilakukan klasifikasi dengan menggunakan data keseluruhan dengan total jumlah data sebanyak 15.228 *tweet*. Hasil dari data keseluruhan yang telah dilakukan tahapan klasifikasi dengan menggunakan metode *SVM* dapat dilihat pada Gambar 3.22.

	tweet	class
0	daftar terima bpnt tahap cair akhir juli cara ...	negatif
1	siapsiap pkh tahap cair akhir juli jelas cara cek	negatif
2	info cair pkh masuk tahap cair pkh tahap juli ...	negatif
3	pkh tahap cair hari begini kategori terima pkh	negatif
4	bansos pkh bpnt cair juli cek syarat salur	negatif
...
15223	sobat suluh dapat berita hoax mohon untuk dapa...	positif
15224	peduli sama bukti cinta kasih	negatif
15225	perintah ikan bantu sosial hadap pandemi covid	positif
15226	salur bantu pangan masyarakat dampak covid des...	negatif
15227	tetap semangat sehat ibu	negatif

15228 rows × 2 columns

Gambar 3. 22 Hasil Data Keseluruhan