



**ANALISIS SENTIMEN PUBLIK TERHADAP PELAYANAN TES
SWAB-PCR COVID-19 DI INDONESIA MENGGUNAKAN
ALGORITMA SUPPORT VECTOR MACHINE**



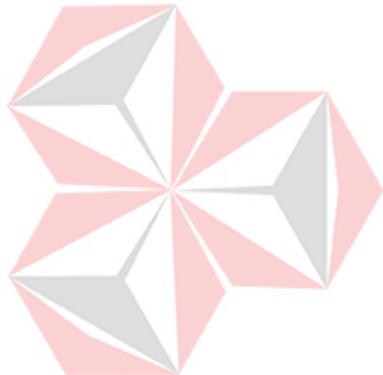
Oleh:
AMIRUL MUKMININ
17410100189

FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA
2021

**ANALISIS SENTIMEN PUBLIK TERHADAP PELAYANAN TES
SWAB-PCR COVID-19 DI INDONESIA MENGGUNAKAN
ALGORITMA *SUPPORT VECTOR MACHINE***

TUGAS AKHIR

**Diajukan sebagai salah syarat untuk menyelesaikan
Program Sarjana**



**UNIVERSITAS
Dinamika**

Oleh:

Nama : Amirul Mukminin
NIM : 17410100189
Program Studi : S1 Sistem Informasi

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA
2021**

Tugas Akhir

ANALISIS SENTIMEN PUBLIK TERHADAP PELAYANAN TES SWAB-PCR COVID-19 DI INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

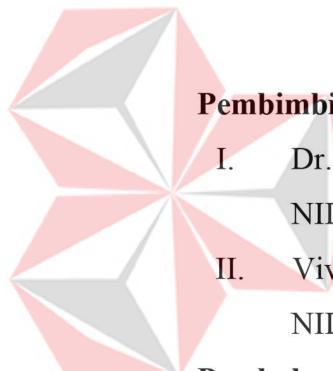
Dipersiapkan dan disusun oleh

Amirul Mukminin

NIM: 17410100189

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahasan

Pada: Senin, 12 Juli 2021



Susunan Dewan Pembahasan

Pembimbing

I. Dr.M.J. Dewiyani Sunarto

NIDN. 0725076301

II. Vivine Nurcahyawati, M.Kom.

NIDN. 0723018101

Pembahasan

Tutut Wurjianto, M.Kom.

NIDN. 0703056702


Digitally signed by
Dewiyani


Digitally signed
by Vivine
Nurcahyawati
Date: 2021.08.10
21:55:01 +07'00'


Digitally signed by
Tutut Wurjianto
Date: 2021.08.12
11:45:55 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana:



Digitally signed by
Universitas Dinamika
Date: 2021.08.12
12:55:48 +07'00'

Tri Sagirani, S.Kom., M.MT.

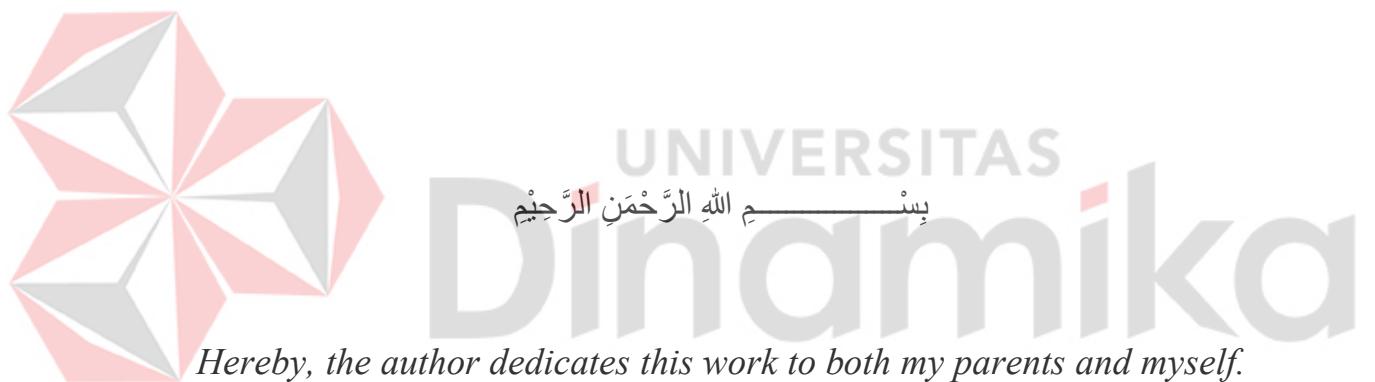
NIDN. 0731017601

Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA

“Do It Now”





SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, saya:

Nama : Amirul Mukminin
NIM : 17410100189
Program Studi : S1 Sistem Informasi
Fakultas : Fakultas Teknologi dan Informatika
Jenis Karya : Tugas Akhir
Judul Karya : **ANALISIS SENTIMEN PUBLIK TERHADAP
PELAYANAN TES SWAB-PCR COVID-19 DI
INDONESIA MENGGUNAKAN ALGORITMA
SUPPORT VECTOR MACHINE**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, dialihmediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 12 Juli 2021

Yang menyatakan



Amirul Mukminin
NIM: 17410100189

ABSTRAK

Adanya layanan swab-PCR juga harus mempertimbangkan berbagai masukan, salah satunya dengan melihat tanggapan publik terhadap pelayanan *swab*-PCR. Berbagai tanggapan terkait pelayanan *swab*-PCR disampaikan masyarakat melalui media sosial Twitter. Melalui pemaparan tersebut, penelitian ini menggunakan data *tweet* berisi tanggapan publik, sehingga menghasilkan klasifikasi dan prediksi yang dapat digunakan sebagai tolak ukur penilaian pelayanan *swab*-PCR. Adanya hal tersebut, penelitian ini akan melakukan analisis sentimen dengan mengklasifikasikan kata data tweet menjadi kata positif dan negatif. Pendekatan yang digunakan ialah algoritma *Support Vector Machine* (SVM). Dengan melalui proses penambangan data dengan RapidMiner, dilanjutkan pelabelan data hasil penambangan, dilanjutkan *text-preprocessing*, kemudian data dibagi menjadi data *training* (80%) dan *testing* (20%), dilanjutkan pembobotan tf-idf di setiap kata pada data berupa nilai numerik, dilanjutkan proses klasifikasi dengan SVM, evaluasi dengan *confusion matrix* dan validasi dengan *cross validation*, serta visualisasi dengan *wordcloud* dan *pie chart*. Sehingga dihasilkan *confusion matrix* untuk validasi dengan perolehan *score* rata-rata 0.66 yang di uji dengan *accuracy* 76%, *precision* 75%, dan *recall* 81%. Berdasarkan *wordcloud* didapatkan empat kata positif ialah gratis, sehat, mandiri, dan positif, serta empat kata negatif ialah kecewa, mahal, tolak, dan antri. Melalui visualisasi *pie chart* menunjukkan bahwa sebesar 45,6% dari total 103 merupakan tanggapan negatif, sedangkan 54,4% merupakan tanggapan positif.

Kata Kunci: analisis sentimen, pelayanan SWAB-PCR, Support Vector Machine, SVM



UNIVERSITAS
Dinamika

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan Tugas Akhir dengan judul “Analisis Sentimen Publik Terhadap Pelayanan Tes Swab-PCR Covid-19 Di Indonesia Menggunakan Algoritma *Support Vector Machine*”.

Penyelesaian Tugas Akhir ini tidak terlepas dari bantuan berbagai pihak yang telah memberikan banyak masukan, nasihat, saran, kritik, dan dukungan moral maupun materil kepada penulis. Oleh karena itu penulis menyampaikan rasa terima kasih kepada:

1. Papa dan Mama tercinta serta keluarga yang selalu mendoakan, mendukung, dan memberikan semangat di setiap langkah dan aktivitas penulis.
2. Bapak Prof. Dr. Budi Jatmiko, M.Pd selaku Rektor Universitas Dinamika Surabaya.
3. Bapak Dr. Anjik Sukmaaji, S.Kom., M.Eng. selaku Ketua Program Studi S1 Sistem Informasi Universitas Dinamika.
4. Ibu Dr.M.J. Dewiyani Sunarto selaku Dosen Pembimbing 1 yang selalu membimbing, mendukung, dan memberikan motivasi kepada penulis dalam menyelesaikan Tugas Akhir ini.
5. Ibu Vivine Nurcahyawati, M.Kom. selaku Dosen Pembimbing 2 yang selalu membimbing, mendukung, memberikan motivasi dan arahan kepada penulis dalam menyelesaikan Tugas Akhir ini.
6. Bapak Tutut Wurjianto, M.Kom. selaku Dosen Pembahas yang telah bersedia menjadi dosen pembahas dalam mengerjakan Tugas Akhir ini.
7. Ibu Tri Sagirani, S.Kom, M.MT. Selaku Dosen Wali yang selalu membimbing dan mendukung dari awal perkuliahan hingga penggerjaan tugas akhir ini
8. Teman-teman tercinta yang memberikan bantuan dan dukungan dalam menyelesaikan Tugas Akhir ini.
9. Pihak-pihak lain yang tidak dapat disebutkan satu-persatu yang telah memberikan bantuan dan dukungan kepada penulis.

Semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah membantu dan memberikan bimbingan serta nasehat dalam proses menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir yang dikerjakan ini masih banyak terdapat kekurangan sehingga kritik yang bersifat membangun dan saran dari semua pihak sangatlah diharapkan agar aplikasi ini dapat diperbaiki menjadi lebih baik lagi. Semoga Tugas Akhir ini dapat diterima dan bermanfaat bagi penulis dan semua pihak.

Surabaya, 12 Juli 2021

Penulis



DAFTAR ISI

Halaman

ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	4
BAB II LANDASAN TEORI.....	5
2.1 Penelitian terdahulu	5
2.2 Analisis Sentimen.....	6
2.3 <i>Text Pre-Processing</i>	7
2.3.1 <i>Case Folding</i>	7
2.3.2 <i>Cleansing</i>	7
2.3.3 <i>Tokenizing</i>	8
2.3.4 <i>Stopword Removal</i>	8
2.3.5 <i>Stemming</i>	9
2.4 Algoritma <i>Support Vector Machine</i> (SVM).....	9
2.5 Twitter.....	11
2.6 Pembobotan <i>Term Frequency-Inverse Document Frequency</i>	12
2.7 <i>Confusion Matrix</i>	13
BAB III METODOLOGI PENELITIAN	14
3.1 Tahap Awal	15
3.1.1 Studi Literatur	15
3.1.2 Teknik Pengumpulan data.....	15

3.2 Tahap Pengembangan Penelitian.....	18
3.2.1 Analisis	18
3.2.2 Penambangan Data	21
3.2.3 Pelabelan	23
3.2.4 <i>Text Pre-Processing</i>	25
3.2.5 Data <i>Training</i> dan Data <i>Testing</i>	27
3.2.6 Pembobotan TF-IDF	28
3.2.7 Klasifikasi <i>Support Vector Machine</i>	29
3.2.8 Evaluasi dan Validasi	31
3.2.9 Visualisasi	32
3.3 Tahap Akhir Penelitian	33
3.3.1 Kesimpulan	33
3.3.2 Saran	33
BAB IV HASIL DAN PEMBAHASAN.....	34
4.1 Analisis	34
4.2 Penambangan Data	34
4.3 Pelabelan	34
4.4 <i>Text Pre-Processing</i>	35
4.5 Data <i>Training</i> dan Data <i>Testing</i>	39
4.6 Pembobotan TF-IDF	40
4.7 Klasifikasi Dengan <i>Algoritma Support Vector Machine</i>	41
4.8 Evaluasi dan Validasi.....	42
4.9 Visualisasi	44
4.9.1 <i>Wordcloud</i>	45
4.9.2 <i>Pie Chart</i>	46
4.9.3 Rekomendasi Hasil Visualisasi	46
4.10Kesimpulan dan Saran	48
BAB V PENUTUP	49
5.1 Kesimpulan	49
5.2 Saran	50
DAFTAR PUSTAKA.....	51
LAMPIRAN	54

DAFTAR TABEL

	Halaman
Tabel 2. 1 Penelitian Terdahulu.....	5
Tabel 2. 2 <i>Case Folding</i>	7
Tabel 2. 3 <i>Cleansing</i>	7
Tabel 2. 4 <i>Tokenizing</i>	8
Tabel 2. 5 <i>Stopword Removal</i>	8
Tabel 2. 6 <i>Stemming</i>	9
Tabel 2. 7 <i>Confusion Matrix</i>	13
Tabel 3. 1 Analisis Kebutuhan Pengguna	19
Tabel 3. 2 Contoh Kamus Positif dan Negatif.....	24
Tabel 3. 3 Contoh Pencocokan Kata Untuk Pelabelan	24
Tabel 3. 4 Contoh Hasil Pembobotan TF-IDF	30
Tabel 3. 5 Hasil Klasifikasi SVM.....	31
Tabel 4. 1 Data Tweet Dengan Unsur Kata Negatif	46
Tabel 4. 2 Data Tweet Dengan Unsur Kata Positif	47
Tabel L1. 1 Contoh Data <i>Tweet</i>	54
Tabel L1. 2 Contoh Proses pembobotan TF-IDF Manual.....	55

DAFTAR GAMBAR

	Halaman
Gambar 2. 2 <i>Hyperplane</i> Pada SVM	9
Gambar 2. 2 Diagram Alir Tahapan SVM.....	11
Gambar 3. 1 Diagram Alir Metodologi Penelitian	14
Gambar 3. 2 Diagram Alir Tahap Awal Penelitian	15
Gambar 3. 3 Studi Dokumentasi.....	16
Gambar 3. 4 Diagram Alir Teknik Pengolahan Data.....	17
Gambar 3. 5 Diagram Alir Tahap Pengembangan Penelitian	18
Gambar 3. 6 Proses <i>Crawling</i> dengan RapidMiner	21
Gambar 3. 7 Parameter <i>Crawling</i>	22
Gambar 3. 8 Menyimpan Data <i>Crawling</i>	23
Gambar 3. 9 Data <i>Tweet</i> Hasil <i>Crawling</i>	23
Gambar 3. 10 Hasil Pelabelan	25
Gambar 3. 11 Diagram Alir Proses <i>Text Pre-processing</i>	26
Gambar 3. 12 Diagram Alir Pembagian Data <i>Training</i> dan <i>testing</i>	28
Gambar 3. 13 Diagram Alir Pembobotan TF-IDF.....	29
Gambar 3. 14 Diagram Alir Proses Klasifikasi SVM.....	30
Gambar 3. 15 Visualisai <i>Hyperplane</i> SVM	31
Gambar 3. 16 Diagram Alir Tahap Akhir Penelitian.....	33
Gambar 4. 1 Data Hasil Pelabelan.....	35
Gambar 4. 2 <i>Upload</i> Data Set ke <i>Drive</i>	35
Gambar 4. 3 Data Set Hasil <i>Import</i>	36
Gambar 4. 4 Data Set Hasil Penggantian Label	36
Gambar 4. 5 Hasil <i>Cleansing</i>	37
Gambar 4. 6 Hasil <i>Case Folding</i>	37
Gambar 4. 7 Hasil <i>Tokenizing</i>	38
Gambar 4. 8 Hasil Penghapusan Tanda Koma.....	38
Gambar 4. 9 Hasil <i>Stopword Removal</i>	38
Gambar 4. 10 Hasil <i>Stemming</i>	39
Gambar 4. 11 Data <i>Training</i>	39

Gambar 4. 12 Data <i>Testing</i>	40
Gambar 4. 13 Hasil Pembobotan TF-IDF	41
Gambar 4. 14 Hasil Uji SVM Pada Data <i>Testing</i>	42
Gambar 4. 15 Hasil <i>10-fold Cross Validation</i>	42
Gambar 4. 16 Rata-rata <i>10-fold Cross Validation</i>	42
Gambar 4. 17 Hasil Prediksi Data <i>Testing</i>	43
Gambar 4. 18 f1, <i>Accuracy</i> , <i>Precision</i> , & <i>Recal Score</i>	43
Gambar 4. 19 Hasil Prediksi <i>Confusion Matrix</i>	44
Gambar 4. 20 Hasil Perubahan Label Menjadi Numerik.....	44
Gambar 4. 21 <i>Wordcloud</i> Sentimen Positif.....	45
Gambar 4. 22 <i>Wordcloud</i> Sentimen Negatif.....	45
Gambar 4. 23 <i>Pie Chart</i>	46
Gambar L2. 1 <i>Code Import</i> Data Set ke <i>Drive</i>	56
Gambar L2. 2 <i>Code</i> Menampilkan Data Berhasil di Import.....	56
Gambar L2. 3 <i>Code</i> Penggantian Label Data Set.....	56
Gambar L2. 4 <i>Code</i> Fungsi Hapus Karakter Dari Twitter.....	57
Gambar L2. 5 <i>Code</i> Fungsi <i>Cleansing</i>	57
Gambar L2. 6 Fungsi <i>Case Folding</i>	58
Gambar L2. 7 <i>Import Library</i> nltk.....	58
Gambar L2. 8 <i>Import</i> Fungsi Tokenizer	58
Gambar L2. 9 Pembuatan Fungsi Tokenizer dalam Variabel "data".....	58
Gambar L2. 10 <i>Code</i> Pemrosesan <i>Tokenizing</i>	59
Gambar L2. 11 Menghapus Tanda Baca Hasil <i>Tokenizing</i>	59
Gambar L2. 12 Install <i>Library</i> Sastrawi	59
Gambar L2. 13 <i>Code Download</i> Stopword dari <i>Library</i> nltk	60
Gambar L2. 14 <i>Code</i> Fungsi <i>Stopword Removal</i>	60
Gambar L3. 1 <i>Code</i> Proses Pembagian Data <i>Training</i> dan Data <i>Testing</i>	61
Gambar L4. 1 <i>Code</i> Pembobotan TF-IDF.....	62
Gambar L5. 1 Menyimpan Data <i>Training</i> Kedalam <i>Drive</i>	62
Gambar L5. 2 <i>Cross Validation</i> dan SVM.....	63
Gambar L6. 1 <i>Import Library</i> sklearn dengan Fungsi cross_val_cross.....	64
Gambar L6. 2 <i>Code Python</i> Menampilkan Hasil <i>10-fold Cross Validation</i>	64



Gambar L6. 3 Menampilkan <i>Average Hasil Cross Validation</i>	64
Gambar L6. 4 <i>Code</i> Proses Prediksi Data <i>testing</i>	65
Gambar L6. 5 <i>Code</i> Menampilkan Hasil Prediksi Data <i>Testing</i>	65
Gambar L6. 6 <i>Import</i> Fungsi metric Untuk Evaluasi <i>Confusion Matrix</i>	66
Gambar L6. 7 Menampilkan Hasil Validasi <i>Confusion Matrix</i>	66
Gambar L6. 8 Inisialisasi Variabel Untuk Menyimpan Hasil Prediksi	66
Gambar L6. 9 <i>Code</i> Untuk Menampilkan <i>Score</i> Hasil Prediksi	67
Gambar L6. 10 Inisialisasi Variabel Data Text dan Label.....	67
Gambar L6. 11 Menggabungkan Data Text dan Data Label Data set.....	68
Gambar L6. 12 Menghapus Kolom yang Tidak Perlu Pada Data Set	68
Gambar L6. 13 Menggabungkan Data Text dan Data Label Data <i>Testing</i>	68
Gambar L6. 14 Menghapus Kolom yang Tidak Perlu Pada Data Set	69
Gambar L7. 2 <i>Import Library</i> nltk dan <i>Download Stopword</i>	69
Gambar L7. 3 <i>Import Stopword List Positive</i>	70
Gambar L7. 4 <i>Word List</i> Positif.....	70
Gambar L7. 5 <i>Import Stopword List Negative</i>	70
Gambar L7. 6 <i>Word List</i> Negatif.....	71
Gambar L7. 7 <i>Import Library Wordcloud</i> dan Matplotlib	71
Gambar L7. 8 Menampilkan <i>Wordcloud</i> Negatif Berdasarkan Label = 0	72
Gambar L7. 9 Menampilkan <i>Wordcloud</i> Positif Berdasarkan Label = 1	72
Gambar L7. 10 <i>Import Library</i> Matplotlib Untuk Visualisasi <i>Pie Chart</i>	73
Gambar L7. 11 Melihat Data Positif dan Negatif untuk persentase <i>Pie Chart</i>	73
Gambar L7. 12 <i>Code</i> Untuk Menampilkahn <i>Pie Chart</i>	73

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Contoh Perhitungan Manual Pembobotan TF-IDF.....	54
Lampiran 2 Proses <i>Text Pre-processing</i> dengan Python	56
Lampiran 3 Proses Pembagian Data (<i>split</i>) dengan Python	61
Lampiran 4 Proses Pembobotan TF-IDF dengan Python	62
Lampiran 5 Proses Klasifikasi SVM dengan Python.....	62
Lampiran 6 Proses Evaluasi dan Validasi dengan Python.....	64
Lampiran 7 Proses Visualisasi dengan Python.....	69
Lampiran 8 Biodata Penulis	74

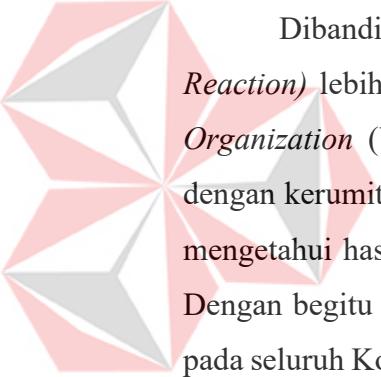


BAB I

PENDAHULUAN

1.1 Latar Belakang

Swab dan PCR tak terpisahkan dalam metode tes untuk menegakkan diagnosis COVID-19. *Swab* adalah suatu cara untuk memperoleh bahan pemeriksaan (sampel). *Swab* dilakukan pada *nasofaring* (hidung) dan atau *orofarings* (tenggorokan) (Makarim, 2020). Pengambilan ini dilakukan dengan cara mengusap rongga hidung dan atau tenggorokan dengan menggunakan alat seperti kapas lidi khusus. Adapun PCR adalah singkatan dari *polymerase chain reaction*. PCR merupakan metode pemeriksaan virus COVID-19 dengan mendeteksi DNA virus. Uji ini akan didapatkan hasil apakah seseorang positif atau tidak (dr. V. Fridawati, 2020).



Dibanding *rapid test*, pemeriksaan RT-PCR (*Real Time - Polymerase Chain Reaction*) lebih akurat. Metode ini (*swab-PCR*) direkomendasikan *World Health Organization* (WHO) untuk mendeteksi COVID-19. Namun akurasi ini diikuti dengan kerumitan proses dan harga alat yang lebih tinggi. Selain itu, proses untuk mengetahui hasilnya lebih lama dibanding *rapid test* (KawalCOVID19.id, 2020). Dengan begitu pemerintah Indonesia berupaya menerapkan pelayanan *swab-PCR* pada seluruh Kota di Indonesia untuk mempercepat pelacakan (*tracing*) penyebaran virus COVID-19. Adanya layanan *swab-PCR* juga harus mempertimbangkan berbagai masukan, salah satunya ialah dengan melihat tanggapan atau opini publik (masyarakat) terhadap pelayanan *swab-PCR*. Hal tersebut sangat perlu dilakukan, karena pemerintah dapat mempertimbangkan pada saat menyikapi tanggapan atau opini publik.

Opini atau tanggapan publik dapat dilihat melalui berbagai media sosial. Media sosial seolah menjadi suatu yang wajib dimiliki oleh setiap orang pada *device* yang dimiliki. Menurut konten HootSuite dan *We Are Social*, dengan laporan yang berjudul “Digital 2021”, pada tahun 2021 Indonesia memiliki populasi penduduk sebesar 274,9 juta jiwa. Dengan 202,6 juta jiwa orang Indonesia di antaranya sudah menjadi pengguna internet. Aktivitas favorit pengguna internet di Indonesia adalah ber-media sosial, jumlah pengguna media sosial di Indonesia

setara dengan 61,8% dari total populasi pada Januari 2021. Sejumlah 170 juta jiwa merupakan pengguna aktif media sosial. Media sosial yang sering digunakan oleh masyarakat Indonesia yaitu salah satunya Twitter. Twitter merupakan media sosial yang bekerja secara *real-time*, yang memungkinkan pengguna untuk mengekspresikan opini dan perasaan mereka mengenai banyak isu atau permasalahan (Pravina, Cholissodin, & Adikara, 2019). Oleh sebab Twitter dapat digunakan sebagai sumber penting opini publik karena berbasis teks untuk bahan penelitian terutama analisis sentimen.

Hal baru yang dapat digunakan sebagai sumber data alternatif ialah pemanfaatan data dari media sosial. Untuk mendapatkan opini atau tanggapan publik, sebagai negara berkembang Indonesia masih mengalami kesulitan akan hal tersebut. Dengan demikian melalui media sosial diyakini dapat digunakan sebagai sumber data yaitu tanggapan atau opini masyarakat yang akan digunakan dalam penelitian ini. Hal-hal atau opini yang masyarakat sampaikan pun belum tentu seluruhnya positif atau negatif. Opini masyarakat mengenai pelayanan *swab-PCR* pada *tweet* perlu dikaji sebagai pemrosesan teks, untuk mengetahui opini tersebut positif atau negatif, sehingga pemerintah dapat menggunakan sebagai salah satu tolak ukur dalam melakukan evaluasi terhadap pelayanan *swab-PCR*. Analisis sentimen merupakan proses yang sangat dibutuhkan dalam menyaring opini-opini masyarakat dan diklasifikasikan ke dalam kelas positif dan negatif (Pravina, Cholissodin, & Adikara, 2019).

Tanggapan atau opini masyarakat (data) yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari Twitter yang dipisahkan menjadi data *training* dan data *testing*. Data *training* sendiri digunakan untuk memperoleh parameter untuk pengklasifikasian dengan algoritma SVM sedangkan *data testing* digunakan untuk melakukan proses evaluasi dari hasil klasifikasi (Taufik & Pamungkas, 2018). SVM mempunyai keunggulan untuk memisahkan data non-linear berdimensi tinggi (Anjasmoros, Istiadi, & Marisa, 2020).

Dengan adanya penelitian ini diharapkan dapat menggunakan data dari *twitter* untuk mengolah tanggapan atau opini masyarakat terhadap pelayanan *swab-PCR*, apakah publik memiliki tanggapan yang positif atau negatif, sehingga dengan diperolehnya hasil klasifikasi tersebut, dapat digunakan sebagai salah satu tolak

ukur pemerintah untuk bahan evaluasi terhadap pelayanan *swab-PCR*. Berdasarkan permasalahan diatas, penelitian ini akan melakukan analisis sentimen untuk melakukan klasifikasi tanggapan atau opini publik terhadap pelayanan *swab-PCR* menggunakan algoritma *Support Vector Machine* (SVM).

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka dapat dirumuskan permasalahan yang ada yaitu: bagaimana melakukan analisis sentimen publik berdasarkan tanggapan atau opini *tweet* pada Twitter terhadap pelayanan *swab-PCR COVID-19* di indonesia menggunakan algoritma *Support Vector Machine*.

1.3 Batasan Masalah

Berdasarkan uraian di atas, untuk menyusun penelitian ini menjadi lebih terarah dan tidak menyimpang dari tujuan pembahasan, maka penulis membatasi pokok permasalahan yang akan dibahas sebagai berikut:

1. Data sekunder yang digunakan dalam penelitian ini berasal dari media sosial yaitu Twitter.
2. Klasifikasi data menggunakan algoritma *Support Vector Machine* (SVM).
3. Proses klasifikasi sentimen dilakukan pada *tweet* yang berbahasa Indonesia.
4. Dalam melakukan ekstraksi data, hanya menggunakan metode *text preprocessing* pada *text mining*.
5. Penelitian ini akan menghasilkan aplikasi pengolahan data yang digunakan sebagai alat bantu *eksperimen* internal peneliti.
6. Data yang digunakan untuk analisis sentimen berasal dari *tweet* pada Twitter dengan *keyword* “pelayanan *swab pcr*”.
7. *Output* dari penelitian ini berupa hasil klasifikasi positif dan negatif, persentase jumlah opini positif dan negatif, serta *wordcloud* yang dapat menampilkan kata yang sering muncul pada data komentar positif atau negatif.

1.4 Tujuan

Berdasarkan latar belakang penelitian ini bertujuan untuk mengetahui tanggapan atau opini publik bersifat positif atau negatif yang berasal dari *tweet* pada media sosial Twitter terhadap pelayanan *swab-PCR* di Indonesia.

1.5 Manfaat

Berdasarkan analisis sentimen yang dilakukan, diharapkan dengan adanya penelitian ini, dapat digunakan untuk mengetahui sentimen publik terhadap pelayanan *swab-PCR* di Indonesia, sehingga hasil sentimen positif atau negatif masyarakat yang berupa opini atau tanggapan dapat dijadikan sebagai salah satu tolak ukur pemerintah untuk bahan evaluasi terhadap pelayanan *swab-PCR*.



BAB II

LANDASAN TEORI

Penelitian ini menggunakan landasan teori yang dijadikan sebagai dasar dalam menyelesaikan masalah yang dijelaskan sebelumnya yaitu penerapan metode *Support Vector Machine* (SVM) yang digunakan dalam analisis sentimen publik terhadap *tweet* pada Twitter. Berikut adalah landasan teori yang digunakan dalam penelitian ini:

2.1 Penelitian terdahulu

Tabel 2. 1 Penelitian Terdahulu

Judul	Penulis	Hasil	Akurasi
Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	Primandani Arsi, Retno Waluyo	Studi kasus : Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma <i>Support Vector Machine</i> (SVM). Tools : RapidMiner <i>Testing</i> : <i>Confusion Matrix</i>	<i>Confusion Matrix</i> accuracy = 96,68%, <i>precision</i> = 95.82%, <i>recall</i> = 94.04% dan AUC = 0,979.
Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode Svm dan Nbc (Studi Kasus: Komentar Pada <i>Play Store</i>)	Mochamad Tri Anjasmoros, Istiadi, dan Fitri Marisa	Studi Kasus : Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode Svm dan Nbc (Studi Kasus: Komentar Pada <i>Play Store</i>) Tools : jupyter notebook (python) <i>Testing</i> : <i>kernel Polynomial</i> dan <i>kernel Sigmoid</i>	Kernel <i>Polynomial</i> : accuracy = 0.5008 <i>precision</i> = 0.5 <i>recall</i> = 1 kernel <i>Sigmoid</i> : accuracy = 0.6244 <i>precision</i> = 0.571 <i>recall</i> = 1
Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme <i>Support Vector Machine</i> .	Arsyia Pravina, Imam Cholissodin, dan Putra Pandu Adikara	Studi kasus : Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme <i>Support Vector Machine</i> . Tools : RapidMiner <i>Testing</i> : <i>Confusion Matrix</i>	<i>Lexicon Bassed Features</i> <i>precision</i> = 40%, <i>recall</i> = 100%, <i>f-measure</i> = 57,14%

- 
1. Berdasarkan penelitian terdahulu yaitu pada tabel 1, 2, dan 3, algoritma SVM telah banyak digunakan untuk penelitian, dan hasil dari akurasinya lebih baik dari pada algoritma *Naive Bayes Clasifier* (NBC) maupun algoritma lainnya. Oleh karena itu, jika digunakan sebagai bahan evaluasi, akan diperoleh hasil yang baik.
 2. Penelitian ini juga menggunakan metode *text mining* yaitu *text preprocessing* dengan beberapa proses yaitu *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*.
 3. Dalam penelitian ini, opini publik yang berasal dari *tweet* pada Twitter akan diolah sehingga menghasilkan klasifikasi positif atau negatif dengan menggunakan algoritma *Support Vector Machine*.
 4. Berdasarkan penelitian terdahulu visualisasi yang digunakan berupa tabel, maka dalam penelitian ini visualisasi akan disajikan dalam bentuk *wordcloud* dan dalam bentuk diagram.
 5. Pada penelitian terdahulu menggunakan *tools* yang berbeda untuk setiap prosesnya, sedangkan penelitian ini menggunakan satu *tools* yang bisa mencakup keseluruhan proses dalam tahap penelitian.

2.2 Analisis Sentimen

Analisis ini digunakan untuk mendapatkan suatu informasi tertentu dari suatu kumpulan data yang ada. Analisis sentimen berfokus pada pengolahan opini yang mengandung polaritas yaitu memiliki nilai sentimen positif atau negatif (Novantirani, Sabariah, & Effendy, 2015). Analisis sentimen, yang disebut juga dengan *opinion mining*, yang merupakan salah satu cabang ilmu dari *text mining*, *natural language program*, dan *artificial intellegence*. Proses yang dilakukan oleh analisis sentimen untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat (Luqyana, Cholissodin, & Perdana, 2018). Selain itu analisis sentimen merupakan bidang ilmu yang menganalisis pendapat, sikap, evaluasi, dan penilaian terhadap suatu peristiwa, topik, organisasi, maupun perseorangan (Liu, 2012).

2.3 Text Pre-Processing

Text mining merupakan sebuah proses penemuan informasi, relasi, dan fakta yang tersembunyi didalam teks ketika dilakukan pemrosesan dan analisis data dalam jumlah besar, struktur yang kompleks dan tidak lengkap, dimensi tinggi, serta data yang terdapat *noise*. *Text mining* juga merupakan subyek penelitian yang sangat baru dan mulai diminati banyak orang.

Dalam penyelesaian masalah, *text mining* biasa digabungkan dengan beberapa subyek lain seperti *Data Mining*, *Natural Language Processing*, dan lain-lain. Dalam *text mining*, terdapat tahap seperti ekstraksi teks menggunakan teknik tertentu, pemrosesan teks atau yang biasa disebut *pre-processing text*, pembobotan atau pemberian indeks pada teks, maupun analisis suatu teks (Pravina, Cholissodin, & Adikara, 2019). Dalam *text pre-processing* terdapat 5 proses di dalamnya, yaitu sebagai berikut :

2.3.1 Case Folding

Case folding merupakan tahapan awal pada *Pre-processing* yang bertujuan untuk mengubah setiap bentuk kata menjadi sama. Hal ini dilakukan dengan mengubah kata menjadi *lower case* atau huruf kecil. Berikut adalah contoh proses *case folding* pada Tabel 2.2

Tabel 2. 2 Case Folding

Tweet Asli (sebelum Case Folding)	Setelah Case Folding
setelah kesana kemari traveling dan harus swab pcr di berbagai tempat. Kadang gw respect sama pelayanan swab pcr yg diberikan ditempat A, kemudian ditempat B beda cuy pelayanannya, kek ada yg kurang gitu dibanding pelayanan yg A tadi.	setelah kesana kemari traveling dan harus swab pcr di berbagai tempat. kadang gw respect sama pelayanan swab pcr yg diberikan ditempat a, kemudian ditempat b beda cuy pelayanannya, kek ada yg kurang gitu dibanding pelayanan yg a tadi.

2.3.2 Cleansing

Cleansing merupakan proses pembersihan kata dengan menghilangkan delimiter koma (,), titik (.), dan tanda baca lainnya, dengan tujuan mengurangi *noise*. Berikut adalah contoh proses *cleansing* pada Tabel 2.3

Tabel 2. 3 Cleansing

Sebelum Cleansing	Setelah Cleansing
setelah kesana kemari traveling dan harus swab pcr di berbagai tempat. kadang gw respect sama pelayanan swab pcr yg diberikan ditempat a, kemudian ditempat b beda cuy	setelah kesana kemari traveling dan harus swab pcr di berbagai tempat kadang gw respect sama pelayanan swab pcr yg diberikan ditempat a kemudian ditempat b beda cuy pelayanannya

Sebelum <i>Cleansing</i>	Setelah <i>Cleansig</i>
pelayanannya, kek ada yg kurang gitu dibanding pelayanan yg a tadi.	kek ada yg kurang gitu dibanding pelayanan yg a tadi

2.3.3 *Tokenizing*

Tokenizing adalah memotong sebuah kalimat berdasarkan tiap kata yang menyusunnya yang dipisahkan dengan tanda koma (,). Di dalam *tokenizing* karakter dan *symbol* selain a-z dihilangkan, pemecahan kalimat dan kata dilakukan berdasarkan pada spasi di dalam kalimat tersebut. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua token ke bentuk huruf kecil. (Indraloka & Santosa, 2017). Berikut adalah contoh proses *Tokenizing* pada Tabel 2.4.

Tabel 2. 4 *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
setelah kesana kemari traveling dan harus swab pcr di berbagai tempat kadang gw respect sama pelayanan swab pcr yg diberikan ditempat a kemudian ditempat b beda cuy pelayanannya kek ada yg kurang gitu dibanding pelayanan yg a tadi	setelah,kesana,kemari,traveling,dan,harus,swab,pcr,di,berbagai,tempat,kadang,gw,respect,sama,pelayanan,swab,pcr,yg,diberikan,ditempat,a,kemudian,ditempat,b,beda,cuy,pelayanannya,kek,ada,yg,kurang,gitu,dibanding,pelayanan,yg,a,tadi

2.3.4 *Stopword Removal*

Stopword merupakan daftar kata umum yang tidak memiliki arti penting dan tidak digunakan. Pada proses ini kata umum akan dihapus untuk mengurangi jumlah kata yang disimpan oleh sistem. Berikut adalah contoh proses *Stopword Removal* pada Tabel 2.5.

Tabel 2. 5 *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
setelah, kesana,kemari,traveling,dan,harus,swab,pcr,di,berbagai,tempat,kadang,gw,respect,sama,pelayanan,swab,pcr,yg,diberikan,ditempat,a,kemudian,ditempat,b,beda,cuy,pelayanannya,kek,ada,yg,kurang,gitu,dibanding,pelayanan,yg,a,tadi	kesana,kemari,traveling,swab,pcr,kadang,respect,pelayanan,swab,pcr,ditempat,beda,pelayanannya,dibanding,pelayanan

2.3.5 Stemming

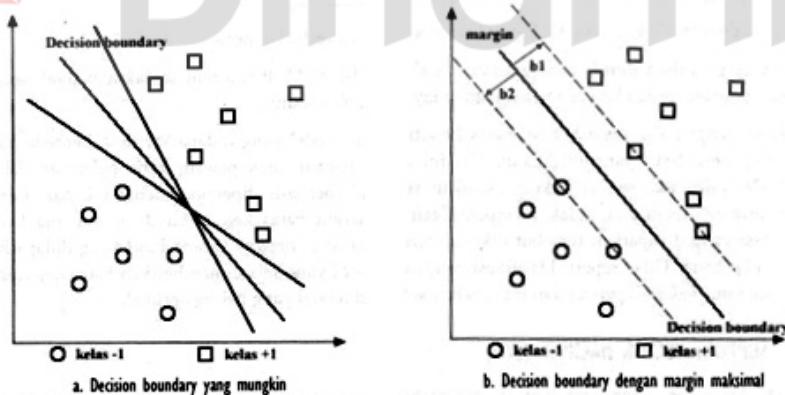
Stemming merupakan proses untuk mencari *stem* (kata dasar) dari kata hasil *stopword removal (filtering)*. Terdapat dua aturan dalam melakukan stemming yaitu dengan pendekatan kamus dan pendekatan aturan (Utomo, 2013). Berikut adalah contoh proses *Stemming* pada Tabel 2.6

Tabel 2. 6 *Stemming*

Sebelum Stemming	Setelah Stemming
kesana,kemari,traveling, <i>swab</i> ,pcr	kesana,kemari,traveling, <i>swab</i> ,pcr
kadang,respect,pelayanan, <i>swab</i> ,pcr	kadang,respect,layanan, <i>swab</i> ,pcr
ditempat,beda,pelayanannya,	tempat,beda,layan,banding,layan
dibanding,pelayanan	

2.4 Algoritma *Support Vector Machine* (SVM)

Metode *Support Vector Machine* (SVM) termasuk kedalam salah satu *machine learning (supervised learning)* yang dapat memprediksi kelas berdasarkan dari hasil proses *training* (pelatihan). Dengan melakukan pelatihan menggunakan data masukan dalam bentuk numerik dan hasil dari ekstraksi fitur (tf-idf) sehingga didapatkan sebuah pola yang akan digunakan untuk pelabelan (Pravina, Cholissodin, & Adikara, 2019).



Gambar 2. 1 *Hyperplane* Pada SVM

Gambar 2.1 menjelaskan konsep klasifikasi SVM. Pada gambar (a) terdapat sejumlah data dengan lingkaran sebagai kelas -1 dan kotak sebagai kelas +1. Pada gambar tersebut juga terdapat sejumlah *hyperplane* yang mungkin untuk set data. Gambar (b) adalah *hyperplane* yang paling maksimal. Perhitungan *hyperplane* dilakukan dengan cara menghitung jarak margin dengan data terdekat dari masing-masing kelas. Data terdekat ini disebut sebagai *Support Vector Machine*. Inti dari

metode ini adalah pencarian *hyperplane* terbaik dari setiap kemungkinan (Neneng, Adi, & Isnanto, 2016). Nilai atau pola dari algoritma SVM adalah sebuah garis pemisah yang disebut dengan *hyperplane*, yang mana garis tersebut berperan dalam memisahkan komentar dengan sentimen positif dengan komentar yang memiliki sentimen negatif, rumus perhitungan *hyperplane* dapat dilihat berikut ini:

Hyperplane klasifikasi linear SVM dinotasikan:

$$f(x) = w \cdot x + b = 0 \quad (1)$$

Dari persamaan di atas didapatkan pertidaksamaan kelas +1(negatif)

$$w \cdot x + b \leq +1 \quad (2)$$

Pertidaksamaan kelas -1:

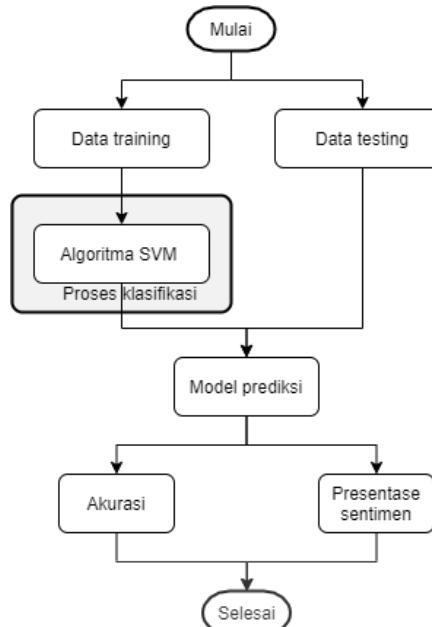
$$w \cdot x + b \geq -1 \quad (3)$$

w adalah bidang normal dan b adalah posisi bidang relatif terhadap koordinat pusat. Dengan mengoptimalkan nilai jarak antara hyperplane dan titik berikutnya, margin terbesar dapat ditemukan, yaitu $1 / \|w\|$. Ini dapat dirumuskan sebagai masalah pemrograman kuadratika (QP) dimana titik minimum persamaan (4) dengan mengingat kendala dari persamaan tersebut(5).

$$\min \frac{1}{2} \|w\|^2 = \min \frac{1}{2} (w_1^2 + w_2^2) \quad (4)$$

$$y_i(w \cdot x_i + b) \geq 1, i = 1, 2, 3, \dots, N \quad (5)$$

Untuk proses dari algoritma SVM dapat dilihat melalui *diagram alir* pada gambar 2.2 berikut ini.



Gambar 2. 2 Diagram Alir Tahapan SVM

2.5 Twitter

Twitter adalah situs *web* yang dimiliki dan dioperasikan oleh Twitter, Inc., yang menawarkan jaringan sosial berupa *microblog*. Disebut *microblog* karena situs ini memungkinkan pengguna mengirim dan membaca pesan *blog* seperti pada umumnya namun terbatas hanya sejumlah 140 karakter yang ditampilkan pada halaman profil pengguna.

Twitter memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus. Pesan dalam Twitter dikenal dengan sebutan *tweet* (Zhang, Ghosh, Dekhil, Hsu, & Liu, 2011). Twitter merupakan media sosial yang memungkinkan pengguna untuk mengekspresikan opini dan perasaan mereka mengenai banyak isu atau permasalahan. Berbeda dengan media sosial yang lain yang harus menjadi teman terlebih dahulu baru dapat berinteraksi, Twitter memungkinkan antarpengguna tetap terhubung walaupun mereka tidak saling berteman.

2.6 Pembobotan *Term Frequency-Inverse Document Frequency*

Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu fitur yang digunakan untuk menghitung pembobotan setiap kata yang telah diekstrak. Dalam TF-IDF, masing-masing kata pada komentar diberikan nilai untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat (Pravina, Cholissodin, & Adikara, 2019).

Penggunaan metode ini umumnya dilakukan untuk menghitung kata umum yang ada pada *information retrieval*. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model *term frequency (tf)* dan *inverse document frequency (idf)*. *Term frequency (tf)* merupakan proses untuk menghitung jumlah kemunculan *term* dalam satu dokumen dan *inverse document frequency (idf)* digunakan untuk menghitung *term* yang muncul di berbagai dokumen yang dianggap sebagai *term* umum, yang dinilai tidak penting (Luqyana, Cholissodin, & Perdana, 2018).

Rumus Tahapan pembobotan dengan TF-IDF adalah sebagai berikut :

1. Hitung *term frequency (tf_{t,d})*
 2. Hitung *weighting term frequency (W_{tf})*
- $$W_{tf} = \begin{cases} t,d & \\ 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$
3. Hitung *document frequency (df)*
 4. Hitung bobot *inverse document frequency (idf)*

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

5. Hitung nilai bobot TF-IDF

$$W_{t,d} = W_{tf_{t,d}} \times idf_t \quad (3)$$

Keterangan :

- $tf_{t,d}$ = Frekuensi *trem*
- $W_{tf_{t,d}}$ = Bobot frekuensi *trem*
- df = Jumlah frekuensi dokumen yang mengandung *trem*
- N = Jumlah total dokumen
- $W_{t,d}$ = Bobot TF-IDF

2.7 Confusion Matrix

Confusion Matrix merupakan teknik yang digunakan untuk mengevaluasi klasifikasi model untuk memperkirakan objek yang benar atau salah. Sebuah matriks dari prediksi akan dibandingkan dengan kelas asli yang berisi informasi *actual* dan prediksi nilai klasifikasi.

Setelah sistem berhasil melakukan klasifikasi pada data *tweet*, dibutuhkan ukuran untuk menentukan seberapa tepat klasifikasi telah dibuat oleh sistem. Tabel 2.7 menunjukkan *confusion matrix* yang digunakan untuk membantu dalam perhitungan sistem evaluasi (Tiara, Sabariah, & Effendy, 2015).

Tabel 2. 7 *Confusion Matrix*

<i>Clasification</i>	<i>Predicted Positives</i>	<i>Predicted Negatives</i>
<i>Actual Positive Cases</i>	<i>Number of True Positive Cases (TP)</i>	<i>Number of False Negative Cases (FN)</i>
<i>Actual Negative Cases</i>	<i>Number of False Positive Cases (FP)</i>	<i>Number of True Negative Cases (TN)</i>

Dalam pengujian *confusion matrix* empat kondisi sebagai berikut: *True Positive*, *True Negative*, *False Positive*, *False Negative*. Jika data positif dan diprediksi positif maka akan dihitung sebagai *True Positive (TP)* dan jika data positif diprediksi negatif maka akan dihitung sebagai *False Negative (FN)*. Jika data negatif diprediksi negatif akan dihitung sebagai *True Negative (TN)* dan jika diprediksi positif maka akan dihitung sebagai *False Positive (FP)* (Flawcett, 2006).

Berdasarkan *matrix confusion*, maka nilai *Accuracy* menggambarkan seberapa akurat model mengklasifikasikan benar dengan rumus sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Precision menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. dengan rumus sebagai berikut:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

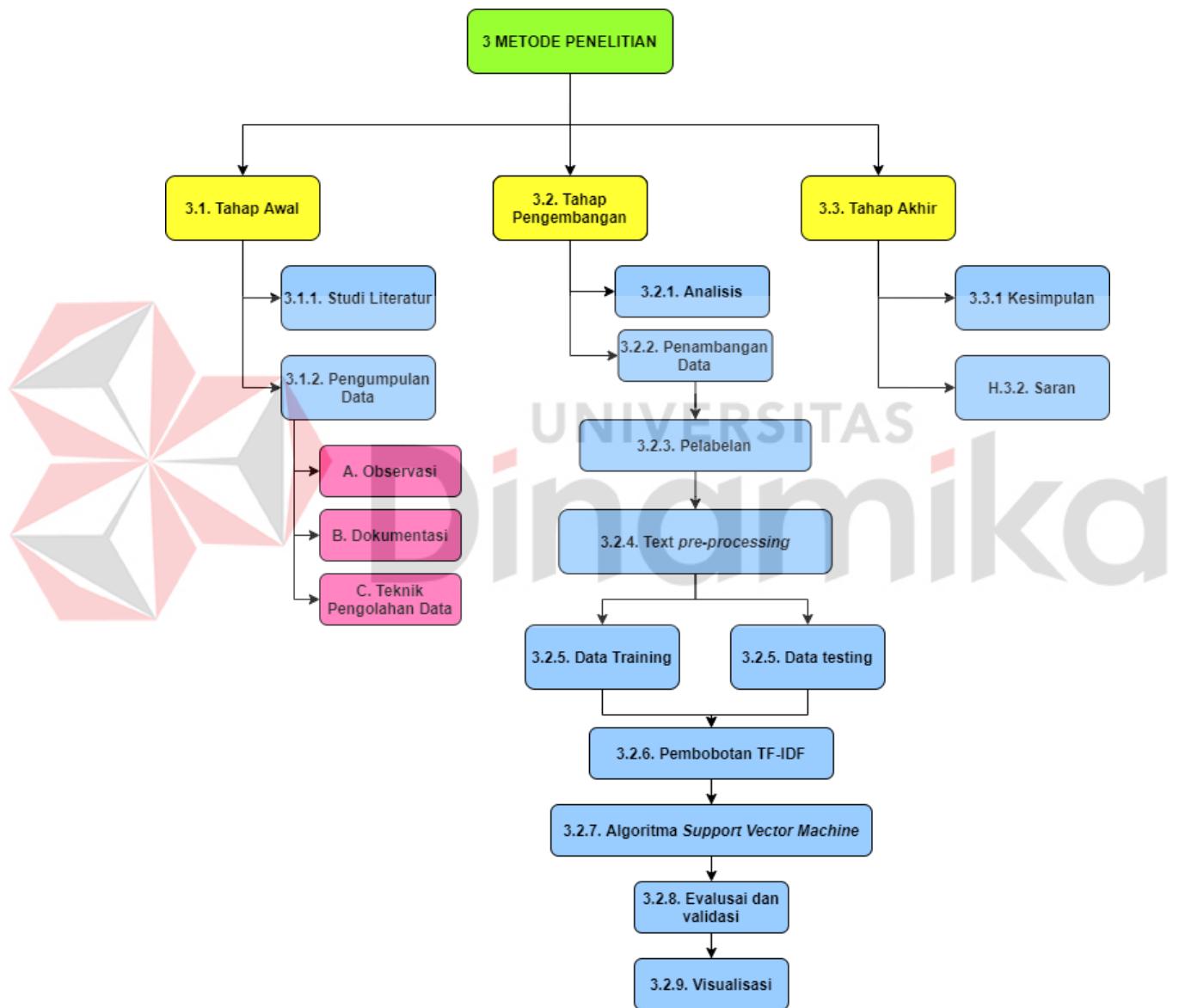
Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Dengan rumus sebagai berikut:

$$\text{Recall} = \frac{TP}{TP + FN}$$

BAB III

METODOLOGI PENELITIAN

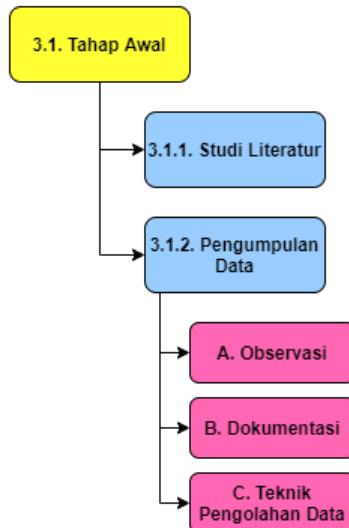
Tahapan penelitian yang digunakan untuk menyelesaikan penelitian ini terdiri dari tahapan awal, pengembangan, dan akhir. Berikut adalah gambaran tahapan penelitian yang dapat dilihat pada Gambar 3.1 berikut ini.



Gambar 3. 1 Diagram Alir Metodologi Penelitian

3.1 Tahap Awal

Hal yang dilakukan dalam tahap awal penelitian dapat dilihat pada gambar 3.2 berikut ini.



Gambar 3. 2 Diagram Alir Tahap Awal Penelitian

Gambar 3.2 merupakan bagian dari tahap penelitian yang penulis lakukan dalam penelitian ini, Adapun rincian jelasnya adalah sebagai berikut:

3.1.1 Studi Literatur

Pada tahapan studi literatur penulis melakukan sebuah kajian sesuai dengan topik yang diambil. Selain itu penulis mencari referensi teori yang membantu dalam menyelesaikan masalah yang telah dijelaskan sebelumnya. Referensi yang dimaksud antara lain penelitian terdahulu, Analisis sentimen. Pelayanan *Swab-PCR* di Indonesia, Algoritma *Support Vector Machine*, *Term Frequency-Inverse Document Frequency* (TF-IDF), dan *Coffusion matrix*.

Referensi yang telah disebutkan di atas dapat ditemukan pada buku, jurnal, artikel laporan penelitian dan situs internet. Referensi ini digunakan sebagai dasar dan acuan teori dalam melakukan studi serta memperkuat permasalahan yang telah dijelaskan sebelumnya.

3.1.2 Teknik Pengumpulan data

Pada tahap pengumpulan data ini dilakukan *crawling* dari Twitter dengan *tweet* berdasarkan *keyword* yang telah ditentukan yaitu “pelayanan swab pcr”.

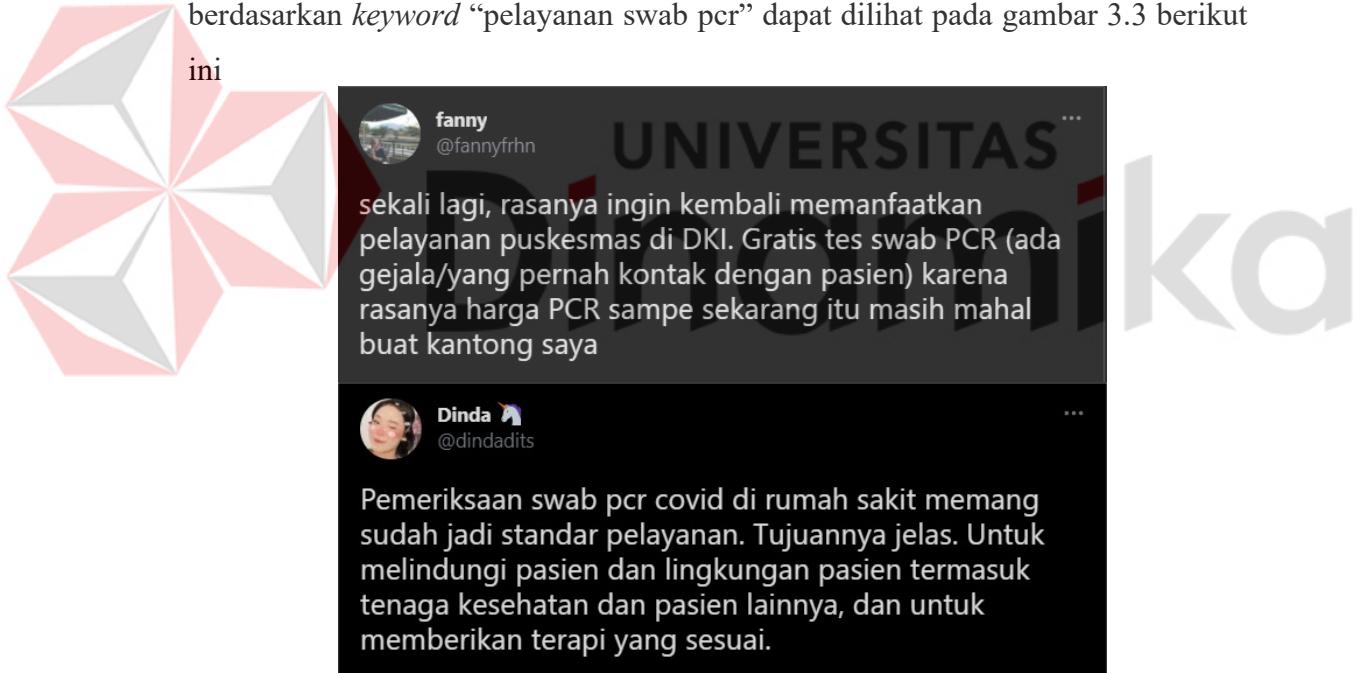
Berikut adalah tahapan dari pengumpulan data yang dilakukan dalam penelitian ini, diantaranya sebagai berikut :

A. Observasi

Observasi yang dilakukan dalam penelitian ini yaitu dengan mengamati dan memahami objek yang akan diteliti, yaitu dengan mengamati *tweet* pada Twitter dengan *keyword* “pelayanan *swab pcr*” yang memuat berbagai macam *tweet* dari berbagai akun pengguna Twitter yang membuat penulis melakukan penelitian terkait dengan *tweet* yang ada pada Twitter.

B. Studi Dokumentasi

Dalam studi dokumentasi penelitian ini, penulis mendapatkan data penelitian ialah tanggapan atau opini berupa *tweet* terhadap pelayanan *swab-PCR* di Indonesia dengan *keyword* “pelayanan *swab pcr*” pada Twitter. Berikut contoh *tweet* berdasarkan *keyword* “pelayanan *swab pcr*” dapat dilihat pada gambar 3.3 berikut ini

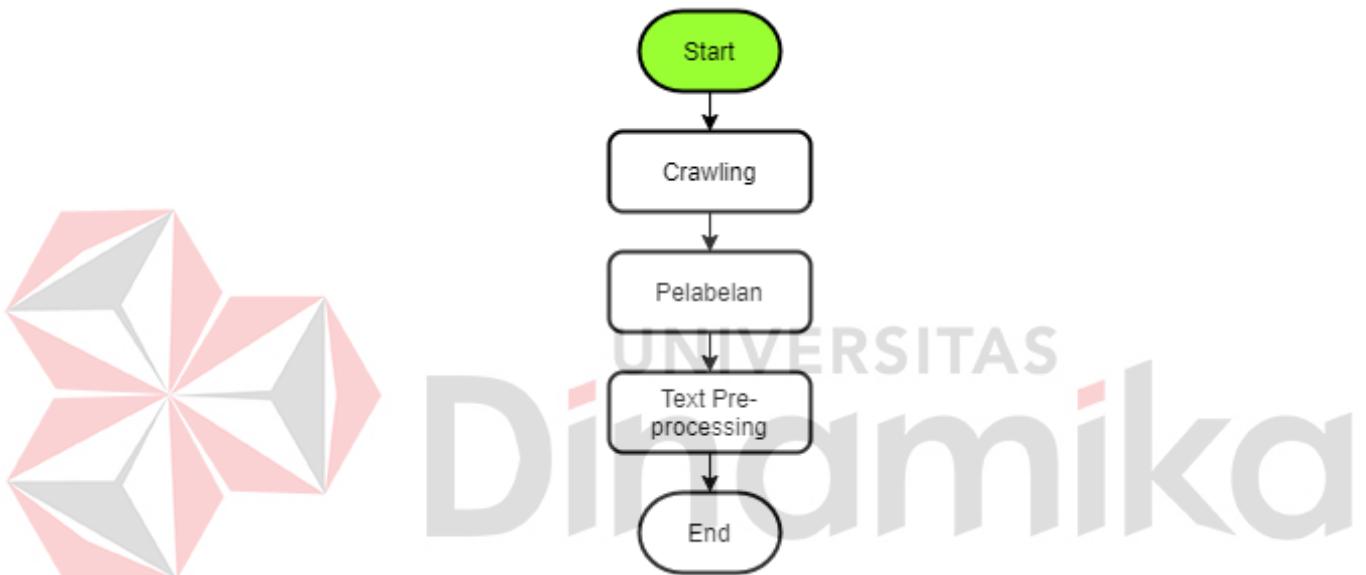


Gambar 3. 3 Studi Dokumentasi

C. Teknik Pengolahan Data

Data awal yang dibutuhkan pada penelitian ini ialah data *tweet* pada Twitter yang akan diambil dengan *crawling* berdasarkan *keyword* yang sudah ditentukan. Kebutuhan data pada penelitian ini ialah data *tweet* yang berisikan opini atau tanggapan publik terkait pelayanan *swab-PCR* di indonesia. Pada proses *crawling* data dilakukan dengan memanfaatkan *tools* RapidMiner. Didalam proses *crawling*

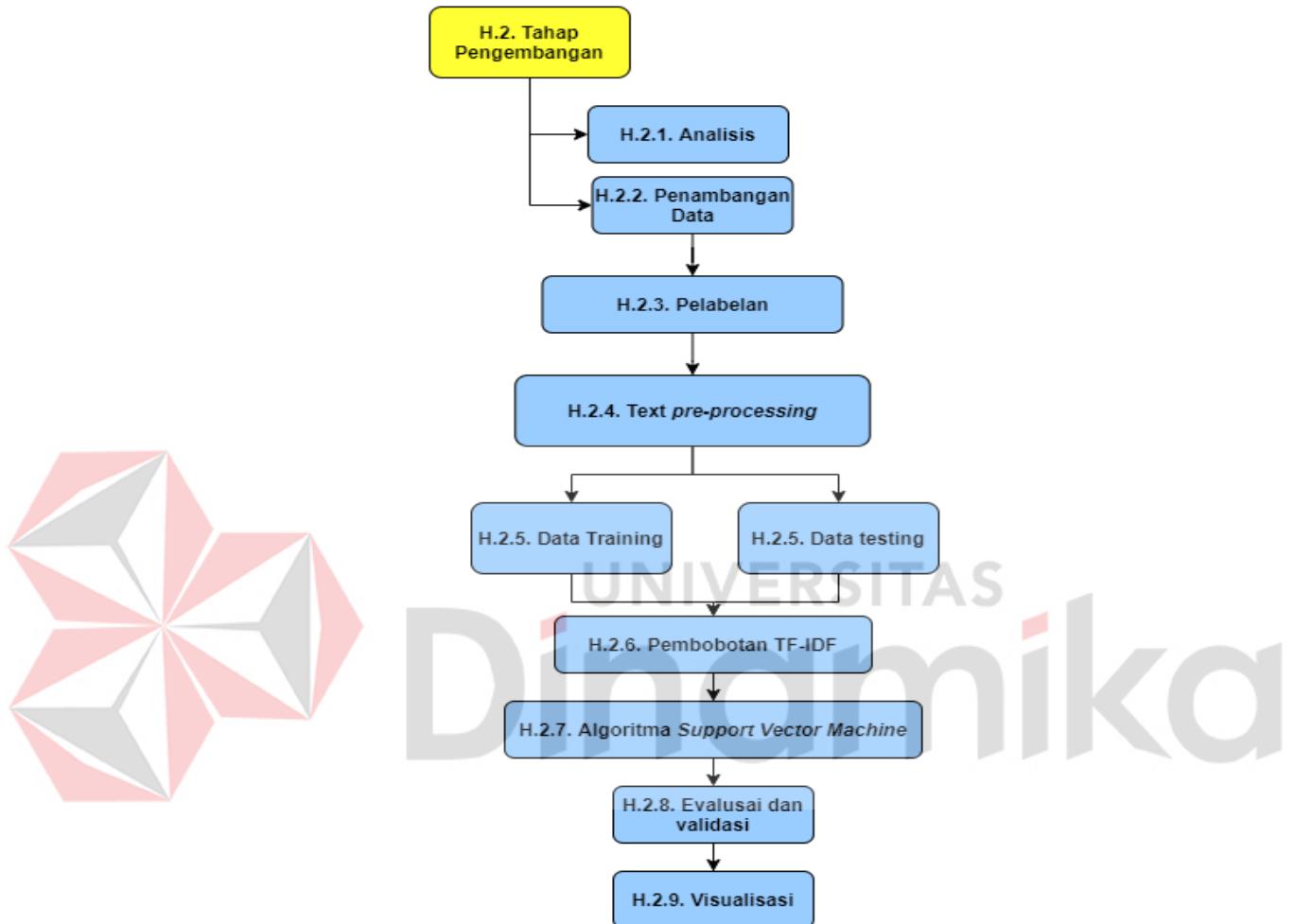
terdapat beberapa parameter yang akan disertakan untuk memperoleh data yang diinginkan guna kebutuhan penelitian. Kemudian data hasil *crawling* akan dilanjutkan dengan pelabelan, dengan dibagi pada 2 kelas yaitu kelas positif dan kelas negatif, dimana proses pelabelan dilakukan berdasarkan setiap kata yang dilakukan oleh sukarelawan. Data hasil *crawling* dapat dikatakan masih belum bersih, dalam artian masih terdapat tanda baca, kata yang tidak penting, tautan, dan lain sebagainya. Oleh karena itu data hasil *crawling* yang sudah berlabel, terlebih dahulu akan dilakukan proses pembersihan yang disebut *text pre-processing*. Diagram alir teknik pengolahan data dapat dilihat pada Gambar 3.4 berikut ini.



Gambar 3. 4 Diagram Alir Teknik Pengolahan Data

3.2 Tahap Pengembangan Penelitian

Bagian ini akan menjelaskan tentang Tahap Pengembangan untuk menyelesaikan penelitian yang digambarkan dengan diagram alir pada Gambar 3.5 di bawah ini.



Gambar 3. 5 Diagram Alir Tahap Pengembangan Penelitian

3.2.1 Analisis

Pada tahap pengembangan hal pertama yang harus dilakukan oleh penulis yaitu melakukan Analisis. Tahap analisis bertujuan untuk mengetahui kebutuhan apa saja yang dibutuhkan dalam penelitian. Pada penelitian ini analisis yang dilakukan terdiri dari lima subbab, yaitu Identifikasi Masalah, Analisis Kebutuhan Pengguna, Analisis Kebutuhan Fungsional, Analisis Kebutuhan Non Fungsional, dan Analisis Kebutuhan Sistem.

A. Identifikasi Masalah

Tahap identifikasi masalah digunakan sebagai dasar dalam melakukan penelitian. Hal ini dilakukan dengan mempelajari studi literatur yang serupa dengan penelitian yang akan dikerjakan penulis, dengan melakukan observasi pada *tweet* di Twitter berbahasa Indonesia dengan *keyword* “pelayanan swab pcr”, sehingga didapatkan informasi yang berkaitan dengan penelitian. Berdasarkan studi literatur, observasi yang sudah dilakukan, digunakan untuk mengetahui komentar/opini publik.

B. Analisis Kebutuhan Pengguna

Analisis Kebutuhan Pengguna dilakukan dengan tujuan untuk mengetahui kebutuhan pengguna yang digunakan atau yang dibutuhkan dalam penelitian ini. berdasarkan hasil Identifikasi Masalah maka dapat dilakukan analisis kebutuhan pengguna untuk penelitian yang akan dibuat, seperti yang sudah dibuat pada Tabel 3.1 berikut ini.

Tabel 3. 1 Analisis Kebutuhan Pengguna

Pengguna	Kebutuhan Pengguna
Peneliti	<ol style="list-style-type: none"> 1. Melakukan survei terhadap <i>tweet</i> pada Twitter berbahasa Indonesia dengan <i>keyword</i> “pelayanan swab pcr” 2. Melakukan pengelompokan komentar 3. Menampilkan hasil dari pengelompokan komentar menurut kelompok positif dan negatif. 4. Menampilkan kata yang sering muncul 5. Manampilkan jumlah kata yang sering muncul

C. Analisis Kebutuhan Fungsional

Analisis Kebutuhan Fungsional dilakukan dengan tujuan untuk mengetahui kebutuhan secara fungsi yang akan digunakan pada penelitian. Berdasarkan kebutuhan pengguna, dibutuhkan kebutuhan fungsional berupa fungsi penambangan data untuk mengumpulkan data, fungsi ekstraksi data, fungsi pengelompokan opini atau pendapat publik menurut kelompok positif atau negatif, dan fungsi visualisasi hasil komentar publik.

D. Analisis Kebutuhan Non Fungsional

Analisis Kebutuhan Non Fungsional dilakukan dengan tujuan untuk mengetahui kebutuhan non fungsi yang akan digunakan. Kebutuhan Non

Fungsional yang dibutuhkan adalah kebutuhan akses dan kebutuhan dalam menampilkan data dalam bentuk visualisasi dengan waktu yang efektif.

E. Analisis Kebutuhan Sistem

1) Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak atau *software requirement* adalah komponen perangkat lunak yang digunakan dalam penelitian ini. Adapun perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut :

Pada penelitian ini *RapidMiner* akan digunakan dalam proses *crawling* atau penambangan data dari Twitter, kemudian Python Google Colab akan digunakan dalam proses *text preprocessing* hingga visualisasi dengan *wordcloud* atau diagram. Adapun *library* yang digunakan untuk mendukung proses pengolahan data yaitu :

- *Library* pandas, *library* ini akan digunakan pada proses *import* data dari csv ke python.
- *Library* NLTK, *Library* ini akan digunakan pada tahap *text preprocessing* yaitu pada proses *tokenizing*.
- *Library* sastrawi, *library* ini akan digunakan pada tahap *text preprocessing* yaitu pada proses *stopword removal* dan juga *stemming*.
- *Library* TfidfVectorizer, *library* ini akan digunakan pada proses perhitungan pembobotan TF-IDF.
- *Library* scikit-learn (sklearn), *library* ini akan digunakan pada proses klasifikasi SVM dan juga validasi dengan *confusion matrix*.
- *Library* wordcloud, *library* ini akan digunakan pada proses visualisasi hasil analisis yaitu berupa kumpulan kata yang sering muncul (*term frequency*).
- *Library* Matplotlib, *library* ini digunakan pada proses visualisasi hasil analisis yaitu berupa grafik persentase.

2) Kebutuhan Perangkat Keras

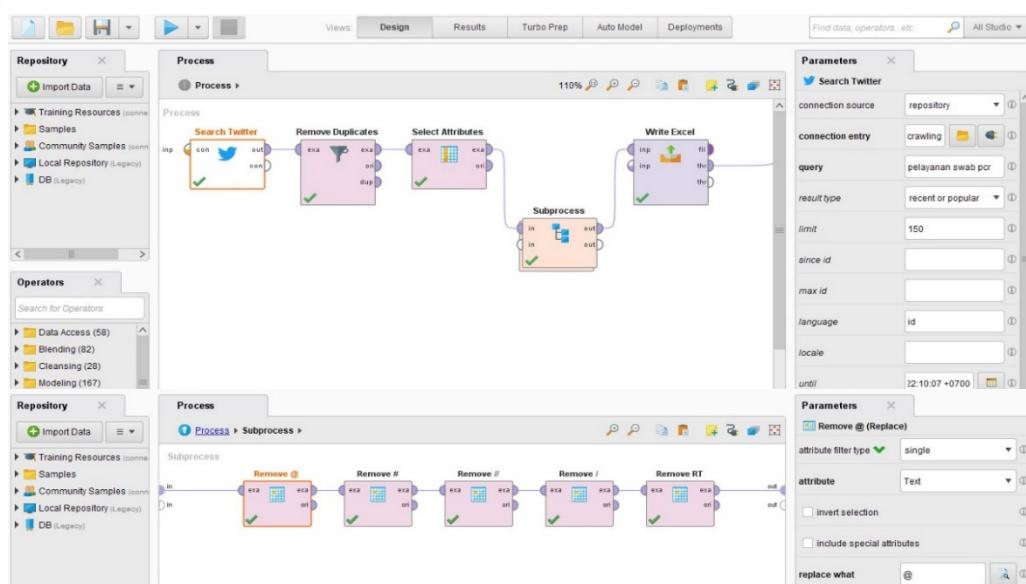
Berdasarkan Analisis Kebutuhan Perangkat Lunak yang telah dibuat kemudian dapat dilanjutkan dengan Analisis Kebutuhan Perangkat Keras atau *Hardware Requirement*. Kebutuhan perangkat keras merupakan peralatan fisik yang membentuk sistem terkomputerisasi terstruktur, serta perangkat keras lain yang

mendukung komputer dalam menjalankan fungsinya. Kebutuhan perangkat keras dibutuhkan ialah sebagai berikut:

- Komputer yang digunakan dalam penelitian ini memiliki *processor* Intel® Core™ i3-4030U.
- Komputer yang digunakan dalam penelitian ini memiliki *Random Access Memory* (RAM) sebesar 6 GB.
- Komputer yang digunakan dalam penelitian ini memiliki monitor dengan resolusi 1366 x 768 pixel.
- Komputer yang digunakan dalam penelitian ini memiliki *Visual Graphic Adaptor* (VGA) NVIDIA GeForce 820M.

3.2.2 Penambangan Data

Tahap kedua ialah melakukan penambangan atau pengambilan data yang akan digunakan dalam penelitian ini. Dalam proses penambangan data digunakan *software* RapidMiner. Hal ini biasa disebut *crawling*, yaitu proses untuk mengambil data *tweet* dari Twitter secara resmi dengan akses API yang diberikan oleh Twitter. Jumlah data yang diambil dari Twitter sejumlah 103 data, yaitu dengan melakukan *crawling* dengan menggunakan RapidMiner. Proses *crawling* dengan *tools* RapidMiner dapat dilihat pada Gambar 3.6 berikut ini.



Gambar 3. 6 Proses *Crawling* dengan RapidMiner

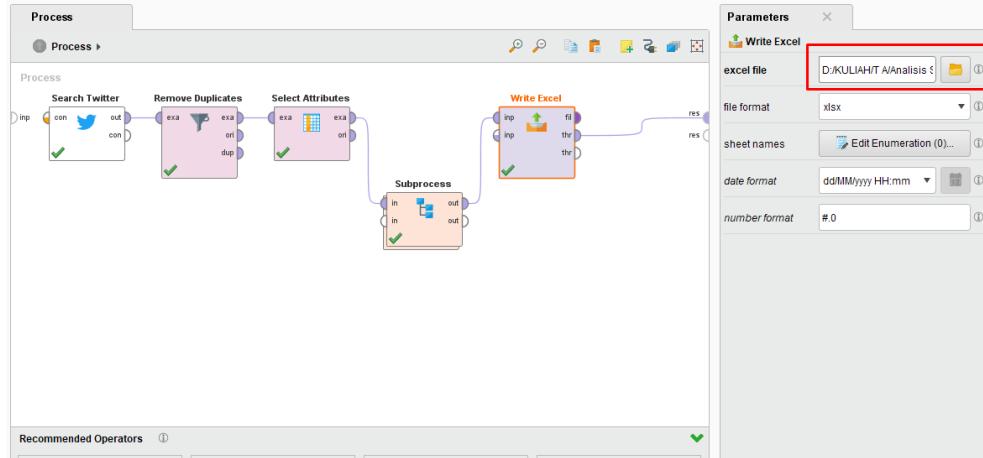
Penambangan data dilakukan dengan melakukan pencarian data (*crawling*) pada Twitter. Tahap *crawling* menggunakan *tools* RapidMiner, yaitu dengan menghubungkannya dengan Twitter menggunakan API Twitter. API diperoleh dengan mengakses Twitter developer yang kemudian dari Twitter developer tersebut akan memberikan *access token*. *Access token* sejumlah 6 digit yang didapatkan akan diinputkan pada RapidMiner. Pada RapidMiner waktu *crawling* dapat ditentukan hingga 7 hari ke belakang yang dapat dikustomisasi pada kolom *until*. Selain waktu *tweet* yang akan diambil diset dengan *keyword* yang bisa disesuaikan melalui kolom *query*, dalam *crawling* ini *query* yang digunakan ialah “pelayanan swab pcr”. Pengambilan data dapat ditentukan dengan *recent or popular*, *recent*, dan *popular*. Selanjutnya batas maksimum data yang akan didapatkan dalam sekali *crawling* dapat ditentukan pada kolom *limit* yang dapat diisi dengan batas maksimal data yang diinginkan. Kemudian untuk menentukan bahasa yang akan diambil pada *tweet* dengan mengisikan kolom *language* dengan “id”, id menginisialisasi bahasa Indonesia. Berikut isian kolom parameter pengambilan data pada Gambar 3.7 berikut ini.



Gambar 3. 7 Parameter *Crawling*

Setelah parameter ditentukan, langkah berikutnya ialah memberikan operator remove duplicate untuk menghapus data yang sama dalam operator ini parameter yang digunakan ialah “single” untuk *attribute filter type*, dan “text” untuk *attribute*. Begitu pula pada operator *Select Attributes* dan juga *Subprocess*. Pada masing-masing *Subprocess* dapat diisikan value berupa karakter yang akan dihilangkan dari

data *crawling*. Tahap terakhir menambahkan operator *Write Excel* untuk menyimpan hasil *crawling* dalam bentuk Excel, pada operator ini nama *file* dan *directory* penyimpanan dapat ditentukan seperti pada gambar 3.8 berikut ini.



Gambar 3. 8 Menyimpan Data *Crawling*

Setelah operator *crawling* tersusun, proses *crawling* dapat dijalankan dengan menekan button *run*. Data hasil *crawling* dapat dilihat pada Gambar 3.9 berikut ini.

No	Text
1	Pelayanan Skrining Sempit test swab antigen, tes swab PCR juga mengambil carian tubuh yang paling banyak mengandung virus, yakni menggunakan alat yang higienis d ¹⁴²¹¹³⁰⁰⁷⁹³⁰⁹⁶⁸⁴⁷³⁸
2	Organisasi Kesehatan Dunia (WHO) pun telah mengumumkan sejak awal pandemik bahanwa tes swab PCR merupakan metode paling akurat untuk mendeteksi Covid-19. te ¹⁴²¹¹³⁰⁰³²⁹²¹¹⁹⁰⁴⁵
3	Kebersihan Laboratorium harus selalu di tinjau, tidak semua Lab bisa menyediakan pelayanan swab-PCR, karena hanya karena debu saja hasil swab PCR bisa terpengaruh ¹⁴²¹¹³²⁸⁰⁵⁰⁵⁰³⁷⁶⁸⁰⁶⁵
4	Jakarta Punya 80 Lab Corona, 10 di Antaranya Gratis,Dinas Kesehatan DKI memastikan 80 lab corona di Jakarta ini sudah sesuai dengan standar yang ditetapkan Kemenkes ¹⁴²¹¹³²⁷⁰³⁵⁴¹⁹⁴⁰²²⁴
5	5 Para NAKES Kerja noni tiap hari, IGD gak pernah sepi, ruangan isolasi ditambah terus untuk meningkatkan pelayanan swab-PCR dan antigen, yang dafar Swab Ag & PCR m ¹⁴²¹¹³¹⁹⁰⁹¹⁸²⁷⁰⁹⁷⁶⁰
6	6 kerja demikian sesuai nasi sekarang sudah tergantikan dengan kerja demikian se swab PCR), pelayanan swab pcr apa sih yg bikin maha.. ¹⁴²¹¹³¹⁷⁷¹⁶⁸⁰⁹⁰³¹⁷¹
7	7 Day 3 PPKM ke Granotic untuk dapat pelayanan swab PCR, udah daftar online buat menghindari antrian, eh masih ngantre ternyata ; ¹⁴²¹¹³¹⁷¹⁹⁰⁰⁸⁸⁴⁵⁶²⁴
8	8 Pelayanan swab pcr sampai saat ini masih diandalkan ketika terdapat hasil rapid atau antigen, untuk dilakukan uji ulang terhadap tes, entah dari segi pelayanannya yg be ¹⁴²¹¹¹⁹⁰³⁵²⁶¹⁵⁵⁸⁷⁹⁰
9	9 di java timur Sejak akhir Desember 2020 hingga awal Januari 2021, LaporCovid-19 menerima total 23 laporan kasus pasien yang ditolak mendapatkan pelayanan swab pc ¹⁴²¹²⁰⁹⁰¹³⁵⁸¹⁹¹³²⁹³⁷
10	10 di tengah musim pandemi ini, pihak yang terkait dengan pelayanan orang masuk orang keluar, untuk bisa memberi bantuan semaksimal mungkin. Kemudian orang yang ¹⁴²¹²⁰⁸⁷⁸⁴⁷⁰³⁷⁷²⁶⁷²¹
11	11 lagi naik lagi, yg mau tracing melalui pelayanan swab pcr, perlu waspada ya..1. Pastikan Semua Alat masih Terbungkus dan Steril.2. Kritis Terhadap Pelayanan yang Diberi ¹⁴²¹²⁰⁸⁶⁶⁰⁵⁰¹²¹¹¹³⁶⁹
12	12 saran aja nh cuy ..pendaftaran pelayanan swab pcr knpa ga dibuat onlie aja ya, biar ga antri gitu.. jadi ketika kita datang langsung tes. ¹⁴²¹²⁰⁸³²⁹⁰⁸⁵³⁴²⁹²⁵³
13	13 pelayanan swab pcr di indonesia semoga semakin baik, bisa tracing makin top ¹⁴²¹²⁰⁸²⁶¹⁴¹⁰²⁴⁶⁶⁵⁶⁰
14	14 Apaan anjinggg lagi dia diputer-puter ya.. "Kalo jan seegini masih pelayanan di puskesmas kelurahan mb"Spmp kelurahan jawabannya ke kecamatan. Capek gak si?dahalan ¹⁴²¹²⁰⁷⁶³¹⁴⁷⁵⁰⁰⁷⁴⁸⁸
15	15 Aneh ini RS pelayanan SWAB PCR, Antigen tampak lebih cepat diurus. Padahal di klinik yang sama ada piasien lain yang akan cek. Nunggunya over timing , & bangsatnya c ¹⁴²¹²⁰³⁷⁰⁷²⁸⁹⁰⁶⁴⁶⁵²⁸
16	16 Selama cukupan pelayanan swab PCR di puskesmas gak sama rata.. jangan harap angka covid akan balikin turun.. screening dan testing yg harusnya diprioritaskan malah dibe ¹⁴¹¹⁹⁷⁵⁵⁴⁵⁵⁷⁵⁸⁰⁹⁰²⁵
17	17 Rockhal-Halo helixlab.id, sudah 36 jam sejak keluaragaanya swab PCR tapi belum ada hasil.. Ini sudah jauh dari janji 1X24 jam.Di whtasapp ke nomer customer service tapi gak ada ¹⁴¹¹⁹⁷³⁵⁶⁸⁵⁵²⁸⁵⁶⁵⁷⁹⁶
18	18 Halo helixlab.id, sudah 36 jam sejak keluaragaanya swab PCR tapi belum ada hasil.. Ini sudah jauh dari janji 1X24 jam.Di whtasapp ke nomer customer service tapi gak ada ¹⁴¹¹⁹⁷³⁴⁶⁴²⁴⁴⁷¹¹⁴²⁴
19	19 sahabatantamedikaRS Antam Medika melayani SWAB PCR dan SWAB Antigen, berikut info Pelayanan Swab PCR Covid-19, mari jaga keseha ¹⁴¹¹⁹⁴³⁰⁸⁰³⁶⁶⁷⁹⁶⁸⁰⁰
20	20 afrikmi Pdh lirs mulai pd colaps ruang ranap & icu full sampe diwilling list, yg kelelahan & tumbang, tpk plasma yg susah dicari, swab pcr yg lama hasilnya emg ga ¹⁴¹¹⁹¹⁷⁶³¹³⁶⁸³³⁶²⁶⁹³
21	21 petugas PKM Kelurahan cuma 8 orang, kerjanya disuruh trading covid, Pd covid, swab PCR, pelayanan berobat umum, pos vaksinasi covid. trus masih dikomplikin karena p ¹⁴¹¹⁵³⁷⁹⁴³⁵³⁴⁰⁴⁷²³⁶
22	22 mozaik atikabuan KavalCOVID19 lya, entre mas blasanya, dan bisanya puskesmas itu ngga punya resources buat bisa pr makar dari itu dialihkan dirujuk ke rspelayar ¹⁴¹¹⁵²⁰⁶¹⁴⁵¹¹¹⁶⁵⁵⁹

Gambar 3. 9 Data Tweet Hasil *Crawling*

Crawling dilakukan 4 kali dalam rentang waktu 7 hari, dari hasil 4 kali *crawling* dapat dikalkulasikan data yang diperoleh 103 data. Contoh data hasil *crawling* dengan RapidMiner dapat dilihat pad Gambar 3.10. Dari hasil *crawling* data disimpan dalam *file* Excel, untuk kemudian akan dilakukan pelabelan.

3.2.3 Pelabelan

Pelabelan atau penentuan kelas atribut diberikan sesuai dengan objektivitas peneliti, pelabelan dilakukan oleh sukarelawan dengan tujuan setiap kata pada data *tweet* bersifat objektif. Pembagian kelas yaitu kelas positif untuk kata pada data

tweet yang bersifat positif yang tersusun menjadi kalimat dan kelas negatif untuk kata pada data *tweet* yang bersifat negatif yang tersusun menjadi kalimat. Pelabelan penentuan kata positif dan negatif ditentukan berdasarkan *load dictionary*, berikut pada Tabel 3.2 adalah contoh kamus dan isinya:

Tabel 3. 2 Contoh Kamus Positif dan Negatif

Kata Positif	Pasti, gratis, upaya, bantu, baik, hebat
Kata Negatif	Korupsi, aneh, jauh, jelek, buruk

Untuk mendapatkan label positif dan negatif dilakukan pencocokan kata pada setiap data dengan kamus kata positif dan negatif, seperti contoh pencocokan kata dari salah satu data tweet pada Tabel 3.3 berikut ini.

Tabel 3. 3 Contoh Pencocokan Kata Untuk Pelabelan

No	Token	Negatif	Positif
1	Setelah	0	0
2	coba	0	0
3	pelayanan	0	0
4	swab	0	0
5	pcr	0	0
6	dirs	0	0
7	hasilnya	0	0
8	positif	0	1
9	tapi	1	0
10	begitu	0	0
11	coba	0	0
12	hasilnya	0	0
13	negatif	1	0
	score	2	1

Dari hasil pencocokan kata pada Tabel 3.3 diketahui kata yang terdaftar dalam kamus mendapatkan nilai 1, sedangkan kata yang tidak terdaftar dalam kamus mendapatkan nilai 0. Dapat dikalkulasikan masing-masing kelas yang mendapatkan nilai 1, dari total 13 kata didapatkan 2 kata negatif dan 1 kata positif dalam satu data *tweet* tersebut, sehingga *score* kata negatif lebih dominan, dari *score* setiap kata dominan tersebut, data *tweet* akan mendapatkan label positif atau negatif. Jika *score* dari kata positif lebih dominan maka data tersebut akan diberikan label sebagai tweet positif, begitu pula sebaliknya. Proses pelabelan data dilakukan dengan melibatkan 4 sukarelawan yang masing-masing sukarelawan akan melabelkan 25% atau sejumlah 25 data *tweet*. Contoh hasil pelabelan data set dapat dilihat pada Gambar 3.10 berikut ini.

comment	value
2. Jika perempuan, yang belum pernah hamil melahirkan. 3. Positif covid-19 yang dibuktikan dengan hasil swab PCR Positif. 4. Negatif covid-19 yang dibuktikan dengan hasil swab PCR negatif atau surat keterangan telah selesai isolasi mandiri dari fasilitas pelayanan kesehatan.	POSITIF
Harusnya bisa vaksin gratis seluruh warga Indonesia tanpa terkecuali, pelayanan gratis di RS, rapidswabpcr pun gratis untuk setiap layanan. Gue ga gerasain tuh 500 triliun dipake buat apa aja, tiap kemana-mana untuk rapidswabpcr gue bayar sendiri, vaksin pun ga semua https://cokatYOmzta	NEGATIF
Setelah coba pelayanan swab pcr di rs A hasilnya positif tapi begitu coba ke rs B hasilnya negatif.. Mirulines	NEGATIF
Kenapa pelayanan swab pcr di indonesia tidak masuk ke desa untuk mempercepat pencegahan penyebaran covid 19??	NEGATIF
[Buka Kerjasama Mitra Sampling Point Swab PCR Test] Bergabung bersama kami menjadi mitra pelayanan test Swab PCR Test bagi rumah sakit, klinik, apotek, praktik dokter mandiri, laboratorium dan sampling point. Dengan harga ekonomis dan hasil cepat (same day service)* https://cozGwG4ZBTB8	POSITIF
Hal Biotizen Budi_waluyo007, Kami melayani juga pelayanan pemeriksaan untuk umum. Pemeriksaan Swab Test PCR tersedia, untuk info lebih lanjut dapat menghubungi kami ni nomor 1-500-810. Tks. BioCare	POSITIF

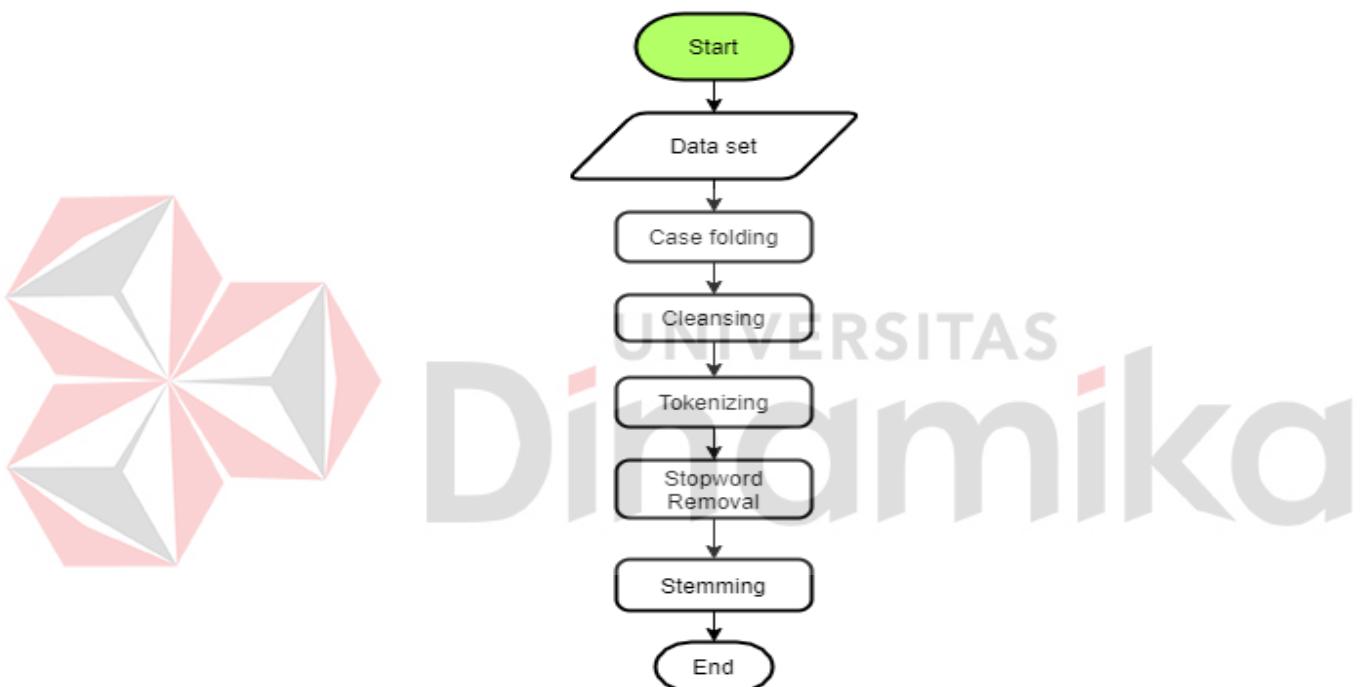
Gambar 3. 10 Hasil Pelabelan

3.2.4 Text Pre-Processing

Berikutnya pada Tahap Pengembangan setelah tahap pelabelan yaitu *Text Preprocessing*. Tahap *text preprocessing* yang dilakukan dalam tahap ini adalah melakukan ekstraksi data. Ekstraksi data bertujuan untuk mengolah hasil penambangan data yang formatnya tidak terstruktur hingga menjadi format data yang tersusun dari kata dasar.

Ekstraksi data yang dilakukan pada penelitian ini terdiri dari lima tahap, yaitu *case folding*, *cleansing tokenizing*, *stopword removal*, dan *stemming*. Dalam program, proses *text pre-processing* diawali dengan penghapusan *link*, *mention*, *retweet*, angka, tanda baca dengan memanfaatkan fungsi “import string, re”, Kemudian dilanjutkan dengan proses *cleansing* pada python yaitu menggunakan fungsi `df_preprocessed`, dimana hasil *cleansing* akan disimpan pada variabel “*tweet*”. *Case folding* pada python menggunakan fungsi `df_preprocessed`, dimana proses *case folding* spesifik pada kolom “Text” sehingga tipe data pada kolom Text akan dijadikan string dengan fungsi “`str`”. Lalu fungsi *tokenizer* dimasukan kedalam variabel yang telah dibuat yaitu “*tokenizer*” yang telah dibuat pada *code* “`def.tokenizing`” dengan parameter “`(data)`”, lalu variabel “*data*” akan digunakan untuk menyimpan hasil *tokenizing*. Selanjutnya pemrosesan *stopword removal* pada python akan dilakukan pada data set. Proses ini menggunakan fungsi “`Sastrawi.StopWordRemover.StopWordRemover`” dengan *library* “`StopwordRemover`”, kemudian variabel “*factory*” digunakan untuk menyimpan *library StopwordRemover*, dan variabel “*stopword*” digunakan untuk menyimpan pembuatan *stopword* yang telah disimpan pada variabel “*factory*”. Proses *stopword*

removal spesifik pada baris “Text” yang dseleksi dengan fungsi “df_preprocessed.iterrows()”, dalam hal ini data hasil *stopword* akan disimpan pada variabel “tweet” dan akan dikembalikan pada variabel “df_preprocessed”. Proses berikutnya ialah *stemming*, dimana proses ini menggunakan fungsi “Sastrawi.Stemmer.StemmerFactory” dengan *library* “StemmerFactory”, kemudian variabel “factory” digunakan untuk menyimpan *library* StemmerFactory, dan variabel “stemmer” digunakan untuk menyimpan pembuatan *stemmer* yang telah disimpan pada variabel factory. Proses *text pre-processing*/ekstraksi data dapat dilihat pada Gambar 3.11 diagram alir berikut ini.



Gambar 3. 11 Diagram Alir Proses *Text Pre-processing*

Dalam *text pre-processing* data set akan masuk dalam tahap *case folding* yang mana data set akan diubah dalam bentuk *lowercase*. kemudian masuk ke tahap *cleansing*, data set akan dibersihkan dengan menghilangkan koma (,), titik (.), dan tanda baca lainnya, dengan tujuan mengurangi *noise*. Selanjutnya pada tahap *tokenizing* data set akan yang awalnya berupa kalimat akan dipisahkan dengan koma (,) sehingga menjadi susunan kata. Kemudian pada tahap *stopword removal*, kata yang tidak memiliki arti akan dihilangkan.

Dan tahap terakhir yaitu stemming, dimana imbuhan pada setiap kata akan dihilangkan sehingga menghasilkan kata dasar. Adapun *library* yang digunakan untuk proses *text pre-processing* ialah sebagai berikut :

A. NLTK (*Natural Language Toolkit*)

Cara kerja *library* NLTK ialah dengan menyediakan alat yang baik dengan mempersiapkan teks sebelum digunakan pada *machine learning* atau algoritma. Dalam *text pre-processing* NLTK dimanfaatkan dalam proses *case folding*, *cleansing* dan juga *tokenizing*.

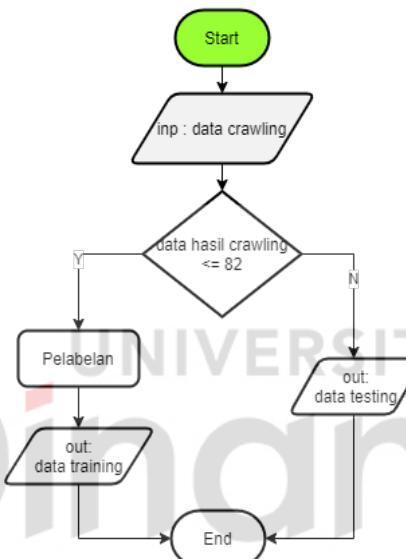
B. Sastrawi

Sastrawi merupakan *library* pada bahasa pemrograman python yang dibangun dengan algoritma NA (Nazief dan Adriani). Algoritma NA adalah aturan yang mengikuti pada aturan Bahasa Indonesia. Aturan tersebut adalah penentuan imbuhan yang di perbolehkan atau tidak. Imbuhan dikelompokkan menjadi imbuhan di awal kata, imbuhan ditengah kata, imbuhan di akhir kata, dan kombinasi. Sastrawi dalam *text pre-processing* dimanfaatkan untuk proses *stopword removal* atau eliminasi kata yang tidak penting, serta *stemming* atau mengembalikan kata dengan imbuhan menjadi kata dasar.

3.2.5 Data Training dan Data Testing

Data *training* dan data *testing* merupakan data set yang telah dibagi dengan perbandingan 80:20 yang akan digunakan dalam penelitian ini. Pembagian porsi 80% untuk data *training*, dan 20% untuk data *testing* yaitu berdasarkan prinsip Pareto, dimana pada prinsip Pareto, jumlah 80:20 menyatakan bahwa 80% hasil, berasal hanya dari 20% usaha yang dilakukan. Sehingga 20% data *training* sudah mewakili 80% data *training* (Loelianto, Sofyan, & Angriani, 2020).

Pada data *training*, data yang digunakan telah berlabel, kemudian dilakukan proses *text pre-processing* untuk dijadikan bahan inputan dari penggunaan algoritma *Support Vector Machine*. Pada bagian proses *testing*, merupakan data yang tidak berlabel atau data yang tidak termasuk data *training*. Meskipun data *testing* tidak diperlukan pelabelan tetapi harus melalui proses *text pre-processing*. Ketika data *testing* melalui proses dalam sistem, yaitu penentuan fitur, mengekstrak fitur dan melakukan klasifikasi sesuai dengan model yang telah ditetapkan, berupa *output* klasifikasi positif dan negatif. Proses pembagian data dapat dilihat pada diagram alir Gambar 3.12 berikut ini.



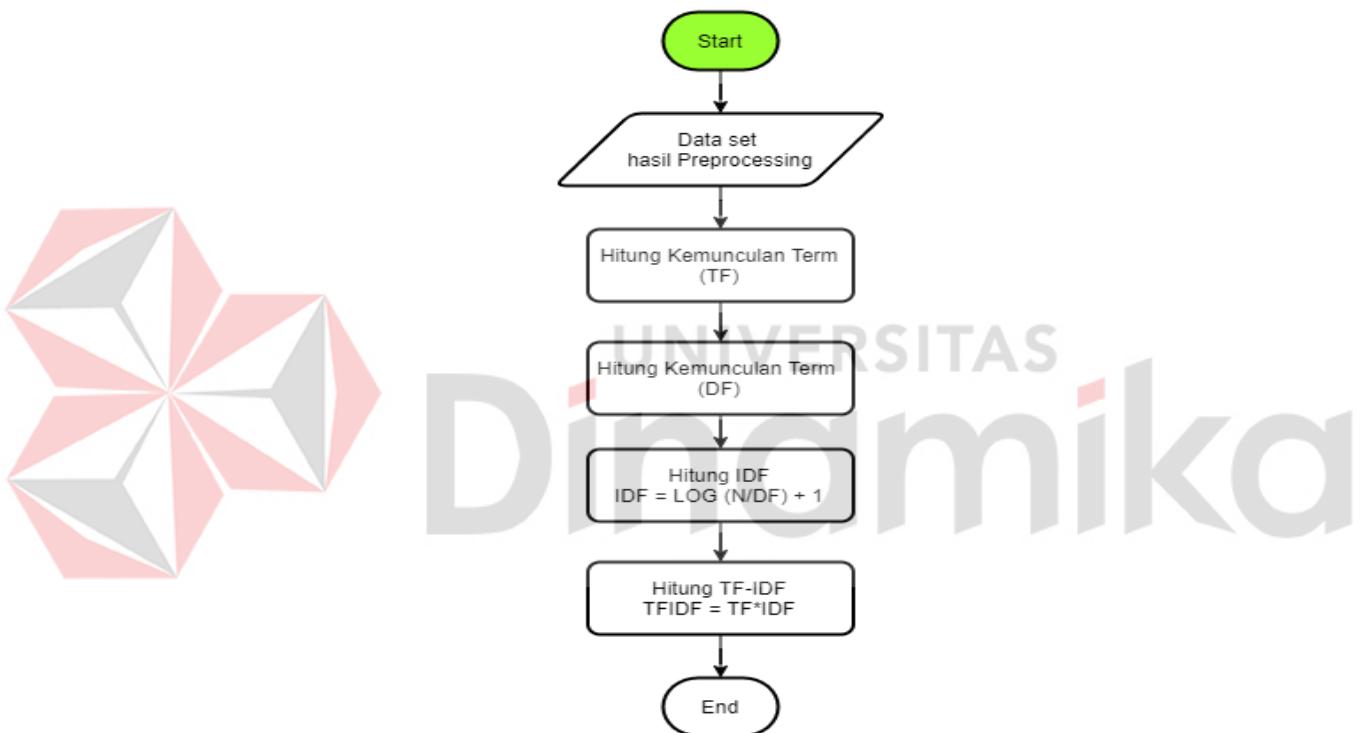
Gambar 3. 12 Diagram Alir Pembagian Data *Training* dan *testing*

Pada gambar 3.12 pembagian dilakukan dengan python dimana data *training* akan diambil sebanyak 80% atau sejumlah 82 data. Jika kondisi tersebut sudah terpenuhi maka sisa data akan dijadikan sebagai data *testing* yaitu sebanyak 20% atau sejumlah 21 data. Proses pembagian data pada python dapat dilihat pada Lampiran 2, serta *code* python dapat dilihat pada Gambar L3.1

3.2.6 Pembobotan TF-IDF

Proses pembobotan TF-IDF pada python akan menggunakan *library* TfidfVectorizer untuk mengubah data yang berupa kata/string menjadi numerik yang disertai dengan bobot pada masing-masing data. TfidfVectorizer adalah algoritma *machine learning* berdasarkan TF-IDF yang khusus mengolah kata-kata dari sebuah dokumen. Maka dengan begitu data numerik lah yang kemudian akan

digunakan dalam proses berikutnya. Dalam hal ini pembobotan dilakukan pada 3 data yaitu data training, testing dan data set, serta digunakan fungsi “fit_transform” pada data training yang berguna untuk menghasilkan parameter model pembelajaran dari data training tersebut, kemudian fungsi “transform” digunakan untuk menghasilkan kumpulan data yang telah diubah (pembobotan). Variabel `x_train` digunakan untuk menyimpan hasil pembobotan data training, variabel `x_test` digunakan untuk menyimpan hasil pembobotan data testing, serta variabel `f_test` digunakan untuk menyimpan hasil pembobotan data set (full data). Diagram alir proses pembobotan TF-IDF dapat dilihat pada Gambar 3.13 berikut ini.

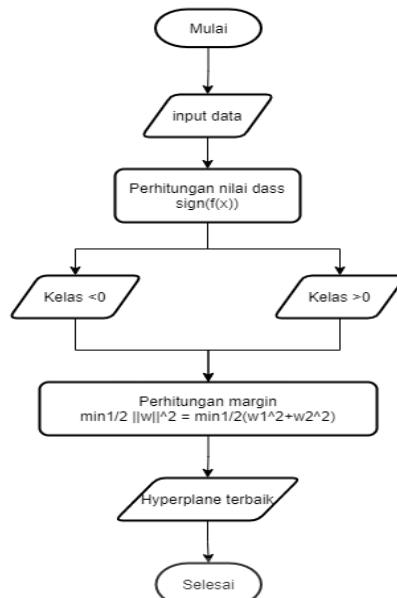


Gambar 3. 13 Diagram Alir Pembobotan TF-IDF

3.2.7 Klasifikasi *Support Vector Machine*

Pada tahap ini akan dilakukan klasifikasi terhadap data yang sudah melewati proses sebelumnya. Dalam tahap pengembangan, algoritma *Support Vector Machine* (SVM) akan diterapkan dalam tahap klasifikasi. Pada python proses klasifikasi dengan SVM memanfaatkan *library* sklearn, dimana sklearn merupakan *library* berbasis python untuk membangun model pembelajaran mesin. *Library* ini juga menyediakan algoritma klasifikasi salah satunya yaitu SVM. Proses klasifikasi akan dilakukan dengan mempelajari data *training* yang telah dibentuk pada subbab

sebelumnya. Berikut tahapan proses kerja SVM dapat dilihat pada Gambar 3.14 berikut ini.



Gambar 3. 14 Diagram Alir Proses Klasifikasi SVM

Dari perhitungan TF-IDF pada Lampiran 1 maka didapatkan data seperti berikut:

Tabel 3. 4 Contoh Hasil Pembobotan TF-IDF

X ₁	X ₂	Kelas (y)
-1	1	-1
1	1	1

- Karena ada dua fitur (X₁ dan X₂), maka bobot (w) juga memiliki dua fitur yaitu (W₁ dan W₂).
- Formulasi yang digunakan adalah sebagai berikut:

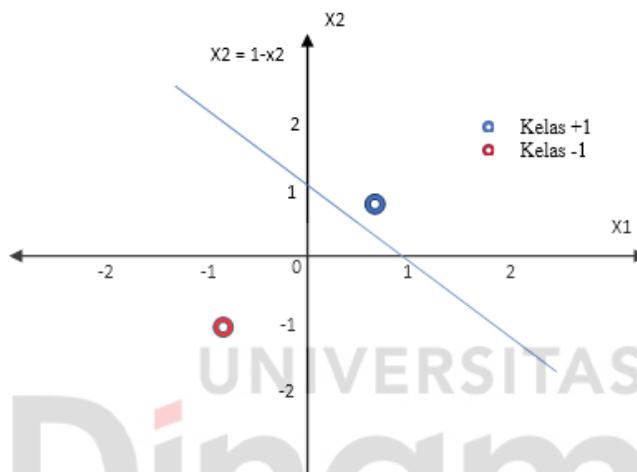
$$\min \frac{1}{2} \|w\|^2$$

$$y_i(w_{xi} + b) \geq 1, i = 1, 2, \dots, N$$

$$y_i(w_1 \cdot x_1 + w_2 \cdot x_2 + b) \geq 1$$

Sehingga didapatkan persamaan *hyperplane*: dengan margin sebagai berikut:

$$\begin{aligned} W_1.X_1 + W_2.X_2 + b = 0 & \quad \min \frac{1}{2} \|w\|^2 = \min \frac{1}{2} (w_1^2 + w_2^2) \\ 1.X_1 + 1.X_2 - 1 = 0 & \quad \frac{1}{2} \|w\|^2 = \frac{1}{2} (1^2 + 1^2) \\ X_1 + X_2 & \quad \frac{1}{2} \|w\|^2 = 2 \\ X_2 = 1 - X_1 & \quad \|w\|^2 = 2 \end{aligned}$$



Gambar 3. 15 Visualisasi *Hyperplane* SVM

Dari hasil visualisasi diatas, objek yang berada diatas garis *hyperplane* masuk kedalam kelas positif, dan yang berada dibawah garis *hyperplane* merupakan kelas negatif. Maka didapatkan hasil seperti pada tabel 3.6 berikut ini.

Tabel 3. 5 Hasil Klasifikasi SVM

		TF-IDF (rank)	SVM (X ₁ , X ₂)	Hyperplane	Hasil
Data 1	sesuai soal swab pcr	2	-1	●	Negatif
Data 2	layan swab pcr kontak positif area antri mesti tunggu banget solusi tracing swab antigen	1	1	○	Positif

3.2.8 Evaluasi dan Validasi

Evaluasi pada penelitian ini dilakukan untuk mengetahui akurasi, dan kualitas klasifikasi terhadap fungsi algoritma *Support Vector Machine*. Kemudian

terdapat validasi yang perlu dilakukan yang bertujuan untuk melihat perbandingan hasil akurasi dari algoritma, metode atau model yang digunakan dengan hasil yang ada. Proses validasi pada penelitian ini menggunakan *K-folds cross-validation* dengan membagi data *training* menjadi beberapa bagian, lalu dengan menggunakan *Confusion Matrix* untuk mengukur akurasi algoritma.

3.2.9 Visualisasi

Untuk menampilkan data hasil analisis dilakukan dengan melakukan visualisasi melalui *wordcloud*, dan diagram lingkaran (*pie chart*). Dengan *wordcloud* akan menampilkan kata yang sering muncul dalam teks yang telah dianalisis sebelumnya, penggunaan ukuran huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data. Kemudian, *pie chart* digunakan untuk menampilkan persentase hasil pengolahan data algoritma SVM terhadap data *tweet* terkait pelayanan *swab-PCR*.

Adapun visualisasi dari penelitian ini, yaitu dengan menyajikan visualisasi dalam 2 bentuk sebagai berikut :

A. *Wordcloud*

Visualisasi data dengan *wordcloud* menghasilkan kata yang sering muncul dalam teks yang telah dianalisis, *library wordcloud* yang digunakan untuk menyajikan visualisasi, dimana dengan visualisasi ini akan membantu para peneliti untuk mendapatkan hasil atau pemahaman secara utuh tentang suatu ide atau gagasan dari seseorang atau tulisan seseorang dalam suatu waktu tertentu. Cara kerja dari *library wordcloud* yaitu dengan memunculkan data teks dengan citra visual yang memberikan kemudahan dalam melakukan interpretasi langsung terhadap ide yang penting. Penggunaan *font-size* huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis serta warna *font* biru dan merah untuk membedakan kelas positif dan negatif.

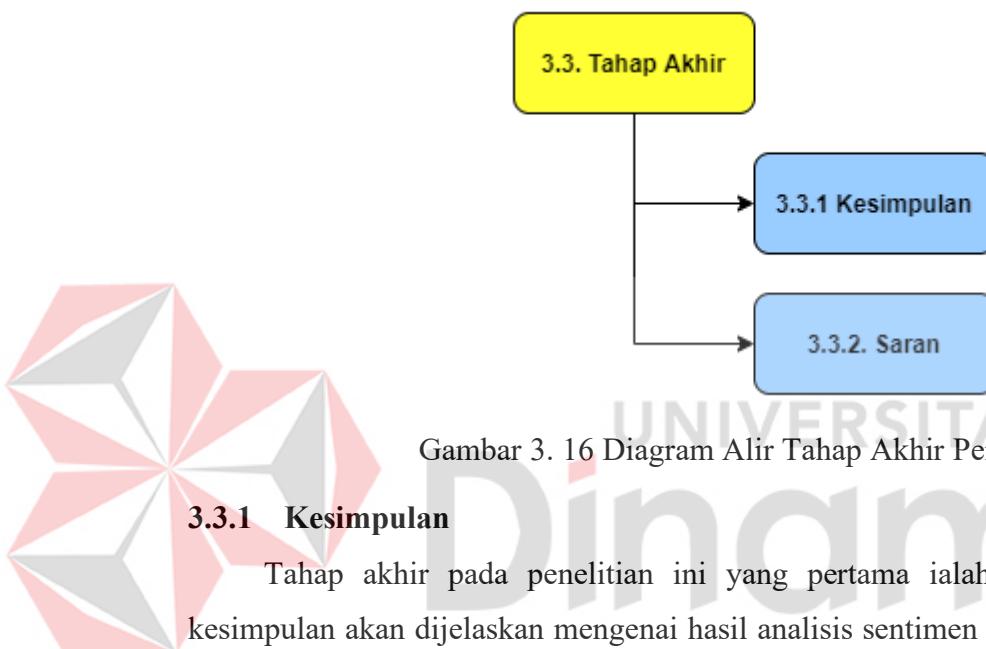
B. *Pie Chart*

Visualisasi data dengan *pie chart* pada python menggunakan *library matplotlib*. Matplotlib merupakan salah satu *library* python yang komprehensif serta berguna untuk menyajikan visualisasi data yang statis dan interaktif dalam berbagai macam pilihan bentuk diagram. Tujuan awal dibuatnya *library* ini adalah untuk membuat

python mampu mereplika kemampuan pembuatan plot. Dalam penelitian ini peneliti memanfaatkan matplotlib untuk menyajikan visualisasi *pie chart*. *Pie chart* digunakan untuk memvisualisasikan hasil persentase SVM dengan membedakannya antara kelas positif dan negatif.

3.3 Tahap Akhir Penelitian

Pada tahap akhir penelitian kesimpulan dari hasil penelitian dan saran untuk penelitian selanjutnya yang lebih baik lagi.



Gambar 3. 16 Diagram Alir Tahap Akhir Penelitian

3.3.1 Kesimpulan

Tahap akhir pada penelitian ini yang pertama ialah kesimpulan. Pada kesimpulan akan dijelaskan mengenai hasil analisis sentimen yang memanfaatkan algoritma SVM untuk mengetahui akurasi klasifikasi dengan *output* berupa klasifikasi data yaitu berupa kelas positif dan negatif berdasarkan data berupa tanggapan atau opini publik terhadap pelayanan tes *swab-PCR COVID-19* dari data *tweet* pada Twitter.

3.3.2 Saran

Saran bertujuan untuk memberi masukan dan sedikit evaluasi mengenai penelitian analisis sentimen tanggapan atau opini publik terhadap pelayanan tes *swab-PCR COVID-19* di Indonesia. Dengan adanya penelitian ini supaya dapat digunakan sebagai bahan penelitian lain dengan topik serupa terkait pelayanan *swab-PCR COVID-19* di Indonesia.

BAB IV

HASIL DAN PEMBAHASAN

Bab keempat pada penelitian ini akan membahas hasil dan pembahasan. Pada bab ini akan dibahas hasil dan implementasi dari penelitian berdasarkan tahapan-tahapan yang sudah dilakukan sesuai dengan metode yang digunakan. Hasil analisis akan dibahas dalam bab ini, berdasarkan pengolahan data *tweet* dari Twitter yang telah di filter berdasarkan *keyword* “pelayanan swab pcr”.

4.1 Analisis

Analisis dimulai dengan tahap penambangan data/*crawling*, *text pre-processing*, membagi data menjadi data *training* dan data *testing*, melakukan perhitungan sentimen manual, pelabelan manual, pengklasifikasian dengan algoritma SVM, dan melakukan visualisasi berupa *wordcloud* dan *pie chart*.

Penelitian ini dilakukan dengan *software* Rapidminer, dan juga Google Colab. Kemudian spesifikasi *hardware* yang digunakan dalam penelitian ini yaitu dengan processor Intel® Core™ i3-4030U dan memory (RAM) sebesar 6GB, *Operating System* Windows 10 Education.

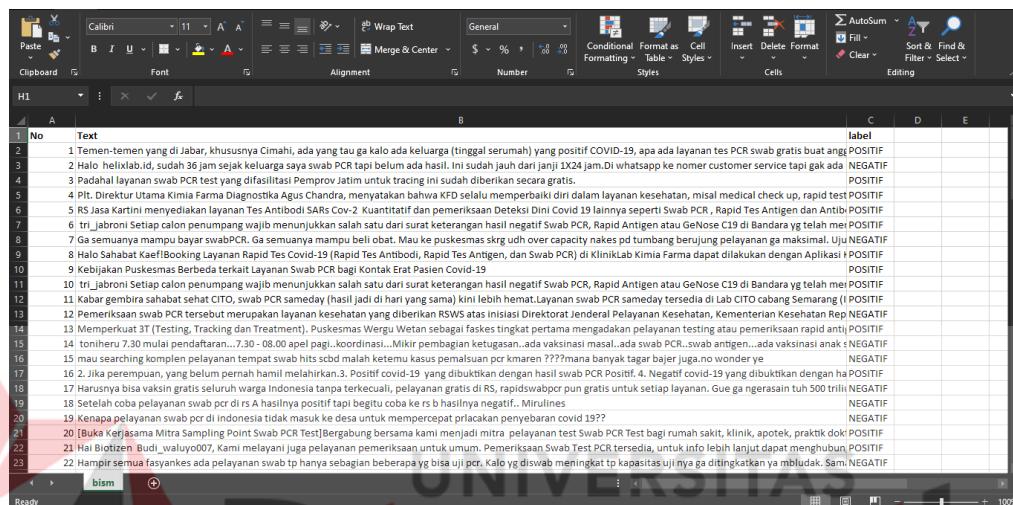
4.2 Penambangan Data

Setelah dilakukannya tahap analisis, akan dilanjutkan pada tahap penambangan data. Pada penelitian ini sangat diperlukan data sebagai pokok bahan analis yang akan dilakukan. Tahap penambangan data sendiri, dalam penelitian ini dilakukan dengan cara *crawling*. Dalam tahap *crawling* menggunakan *tools* RapidMiner, yang kemudian data hasil *crawling* akan disimpan dalam bentuk ekstensi excel (.xls).

4.3 Pelabelan

Setelah didapatkanya data dari hasil *crawling* maka kemudian akan dilakukan pelabelan, pelabelan dilakukan pada data hasil *crawling*. Pelabelan dilakukan pada 103 data yang didapatkan dari hasil empat kali *crawling* pada media sosial Twitter dengan RapidMiner. Setiap sekali *crawling* akan mendapatkan 25 sampai dengan 35 data *tweet* yang mengandung *keyword*. Pelabelan akan dilakukan oleh 4 sukarelawan dan masing-masing sukarelawan akan melabelkan 25 oleh 1

sukarelawan dan 26 data oleh 3 sukarelawan lainnya, dan memberikan kelas pada masing-masing data yaitu kelas positif dan negatif. Proses pemberian label dilakukan berdasarkan pencocokan dari setiap kata penyusun kalimat *tweet* dengan kamus kata positif dan negatif sehingga, jika kata penyusun kalimat tweet terdapat kata positif maka data akan mendapatkan label positif, begitu juga sebaliknya. Untuk contoh beberapa data yang sudah dilakukan pelabelan oleh sukarelawan, dapat dilihat pada Gambar 4.1 berikut ini.

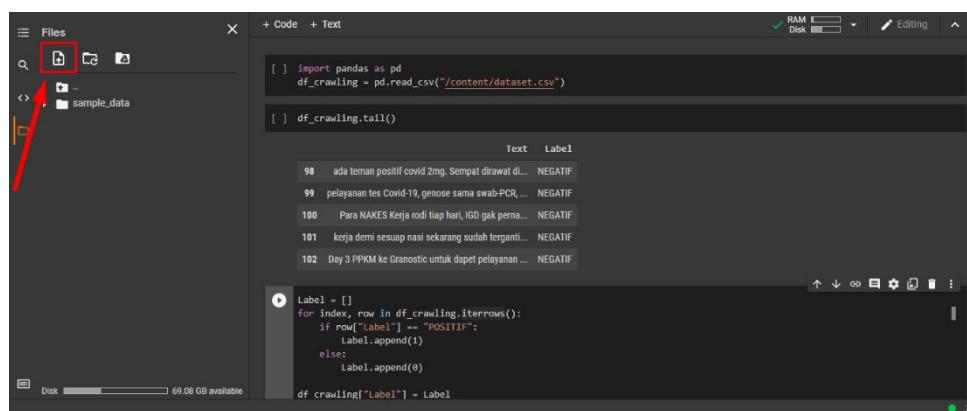


No	Text	label
1	Temen-temen yang di Jabar, khususnya Cimahi, ada yang tau ga kalo ada keluarga (tinggal serumah) yang positif COVID-19, apa ada layanan tes PCR swab gratis buat anggota keluarga?	POSITIF
2	Halo helixlab.id, sudah 36 jam sejak keluarga saya swab PCR tapi belum ada hasil. Ini sudah jauh dari janji 1X24 jam. Di whatsapp ke nomer customer service tapi gak ada	NEGATIF
3	Padahal layanan swab PCR test yang difasilitasi Pemprov jatim untuk tracing ini sudah diberikan secara gratis.	POSITIF
4	Pt. Direktur Utama Kimia Farma Diagnostika Agus Chandra, menyatakan bahwa KFD selalu memperbaiki diri dalam layanan kesehatan, misal medical check up, rapid test	POSITIF
5	RS Jasa Kartini menyediakan layanan Tes Antibodi SARS Cov-2. Kuantitatif dan pemeriksaan Deteksi Dini Covid 19 lainnya seperti Swab PCR, Rapid Tes Antigen dan Antib	POSITIF
6	tr_jabroni Setiap calon penumpang wajib menunjukkan salah satu dari surat keterangan hasil negatif Swab PCR, Rapid Antigen atau GeNose C19 di Bandara yg telah mei	POSITIF
7	Ga semuanya mampu bayar swabPCR. Ga semuanya mampu beli obat. Mau pakeusmane skrg udah over capacity nakes pd tumbang berurung pelayanan ga maksimal. Uju NEGATIF	
8	Halo Sahabat KaelBooking Layanan Rapid Tes Covid-19 (Rapid Tes Antibodi, Rapid Tes Antigen, dan Swab PCR) di KlinikLab Kimia Farma dapat dilakukan dengan Aplikasi i	POSITIF
9	Kebijakan Puskesmas Berbeda terkait Layanan Swab PCR bagi Kontak Erat Pasien Covid-19	POSITIF
10	tri_jabroni Setiap calon penumpang wajib menunjukkan salah satu dari surat keterangan hasil negatif Swab PCR, Rapid Antigen atau GeNose C19 di Bandara yg telah mei	POSITIF
11	Kabar gembira sahabat sehnia CITO, swab PCR sameday (hasil jadi di hari yang sama) kini lebih hemat. Layanan swab PCR sameday tersedia di Lab CITO cabang Semarang (I	POSITIF
12	Pemeriksaan swab PCR tersebut merupakan layanan kesehatan yang diberikan RSWs atas instansi Direktorat Jenderal Pelayanan Kesehatan, Kementerian Kesehatan Rep	NEGATIF
13	Mempertakut 3T (Testing, Tracking dan Treatment). Puskesmas Wergu Wetan sebagai faskes tingkat pertama mengadakan pelayanan testing atau pemeriksaan rapid anti	POSITIF
14	tonihetu 7.30 mulai pendaftaran...7.30 - 08.00 apel pagi...koordinasi...Mikir pembagian ketugasan..ada vaksinasi masal..ada swab PCR..swab antigen...ada vaksinasi anak	NEGATIF
15	mau searching komplen pelayanan tempat swab hits sdcal malam ketemu kasus pemalsuan prk kmaren????mana banyak tagar baper juga,no wonder yg	NEGATIF
16	16.2. Jika perempuan, yang belum pernah hamil melahirkan.3. Positif covid-19 yang dibuktikan dengan hasil swab PCR Positif. 4. Negatif covid-19 yang dibuktikan dengan ha	POSITIF
17	17. Harusnya bisa vaksin gratis seluruh warga Indonesia tanpa terkecuali, pelayanan gratis di RS, rapidswabpcr pun gratis untuk setiap layanan. Gue ga ngerasain tuh 500 trilli NEGATIF	
18	18. Setelah coba pelayanan swab pcx di rs A hasilnya positif tapi begini coba ke rs B hasilnya negatif.. Mirulines	NEGATIF
19	19. Kenapa pelayanan swab pcx di indonesia tidak masuk ke desa untuk mempercepat plafakan penyebaran covid 19???	NEGATIF
20	[BuKerjasma Mitra Sampling Point Swab PCR Test]Bergabung bersama kami menjadi mitra pelayanan test Swab PCR Test bagi rumah sakit, klinik, apotek, praktik dokter	POSITIF
21	21. Hai Biotzen Budi_waluyo007, Kami melayani juga pelayanan pemeriksaan untuk umum. Pemeriksaan Swab Test PCR tersedia, untuk info lebih lanjut dapat menghubungi	POSITIF
22	22. Hampir semua fasyankes ada pelayanan swab tp hanya sebagian beberapa yg bisa uji pr. Kalu yg diswab meningkat tp kapasitas uji nya ga ditingkatkan ya mbiludak. Sami	NEGATIF
23		

Gambar 4. 1 Data Hasil Pelabelan

4.4 Text Pre-Processing

Untuk melakukan proses Text-preprocessing pada python dat set akan di *upload* terlebih dahulu kedalam *cloud drive* Google Colab dengan cara berikut ini. Untuk *upload* data set, seperti pada Gambar 4.2 klik pada button *upload* yang ditunjukan tanda panah warna merah, kemudian pilih file data set.



```

[ ] import pandas as pd
df_crawling = pd.read_csv("/content/dataset.csv")

[ ] df_crawling.tail()

```

Text	Label
98 ada teman positif covid 2mg. Sempat dirawat di...	NEGATIF
99 pelayanan tes Covid-19, genesen sama swab-PCR,...	NEGATIF
100 Para NAKES Kerja rodil tiap hari, IGD gak perna...	NEGATIF
101 kerja demi sesuap nasi sekarang sudah tergant...	NEGATIF
102 Day 3 PPKM ke Granotic untuk dapat pelayanan ...	NEGATIF

```

Label = []
for index, row in df_crawling.iterrows():
    if row["Label"] == "POSITIF":
        Label.append(1)
    else:
        Label.append(0)

df_crawling["Label"] = Label

```

Gambar 4. 2 Upload Data Set ke Drive

Kemudian untuk melakukan *import* data set ke dalam program python, dalam hal ini memanfaatkan *library* pandas untuk *code import* dataset dapat dilihat pada Lampiran 2 Gambar L2.1.

Dari hasil *code* pada Gambar L2.2, akan menampilkan data set dengan jumlah baris yang sesuai. Fungsi tail berguna untuk menampilkan 5 data baris terakhir pada data set. Pada Gambar 4.3 yang menampilkan data set, baris terakhir yaitu 102 yang menunjukan total keseluruhan dat set berjumlah 103 dengan urutan yang dimulai dari angka 0.

	Text	Label
98	ada teman positif covid 2mg. Sempat dirawat di...	NEGATIF
99	pelayanan tes Covid-19, genose sama swab-PCR, ...	NEGATIF
100	Para NAKES Kerja rodii tiap hari, IGD gak perna...	NEGATIF
101	kerja demi sesuap nasi sekarang sudah tergant... 102 Day 3 PPKM ke Granostic untuk dapet pelayanan ...	NEGATIF

Gambar 4. 3 Data Set Hasil *Import*

Setelah data di *import*, label pada data set yang awalnya berupa “POSITIF” dan “NEGATIF” akan diganti menjadi “0” dan “1”. Penggantian label bertujuan untuk memudahkan proses inisialisasi label pada tahap klasifikasi. Dengan *code* pada Gambar L1.3. Sehingga didapatkan hasil penggantian label seperti pada Gambar 4.4 berikut ini. dimana pada kolom label yang awalnya berupa POSITIF atau NEGATIF akan menjadi 0 untuk label NEGATIF dan 1 untuk label POSITIF.

	Text	Label
0	Temen-temen yang di Jabar, khususnya Cimahi, a...	1
1	Halo helixlab.id, sudah 36 jam sejak keluarga...	0
2	Padahal layanan swab PCR test yang difasilitas...	1
3	Pit. Direktur Utama Kimia Farma Diagnostika Ag...	1
4	RS Jasa Kartini menyediakan layanan Tes Antibio...	1
...
98	ada teman positif covid 2mg. Sempat dirawat di...	0
99	pelayanan tes Covid-19, genose sama swab-PCR, ...	0
100	Para NAKES Kerja rodii tiap hari, IGD gak perna...	0
101	kerja demi sesuap nasi sekarang sudah tergant... 102 Day 3 PPKM ke Granostic untuk dapet pelayanan ...	0
103 rows x 2 columns		

Gambar 4. 4 Data Set Hasil Penggantian Label

Selanjutnya masuk ke dalam proses *text pre-processing*. Dalam program, proses *text pre-processing* diawali dengan penghapusan atribut yang tidak diperlukan yang berasal dari Twitter seperti *emoticon*, *link*, dan sebagainya, untuk *code* tahap ini dapat dilihat pada Gambar L1.4.

Dari hasil *code* fungsi cleansing yaitu *code* pada Gambar L1.5 maka akan dihasilkan data *cleansing* atau penghilangan karakter seperti pada Gambar berikut.

	Text	Label
0	Temen temen yang di Jabar khususnya Cimahi a...	1
1	Halo helixlab id sudah jam sejak keluarga s...	0
2	Padahal layanan swab PCR test yang difasilitas...	1
3	Plt Direktur Utama Kimia Farma Diagnostika Ag...	1

Gambar 4. 5 Hasil *Cleansing*

Setelah *cleansing*, selanjutnya akan dilakukan tahap *case folding*. *Code* untuk fungsi *case folding* dapat dilihat pada Gambar L1.6. Dari hasil *case folding* seluruh data pada kolom “Text” akan berubah menjadi *lowercase* atau huruf kecil semua, seperti pada Gambar berikut ini.

	Text	Label
0	temen temen yang di jabar khususnya cimahi a...	1
1	halo helixlab id sudah jam sejak keluarga s...	0
2	padahal layanan swab pcr test yang difasilitas...	1

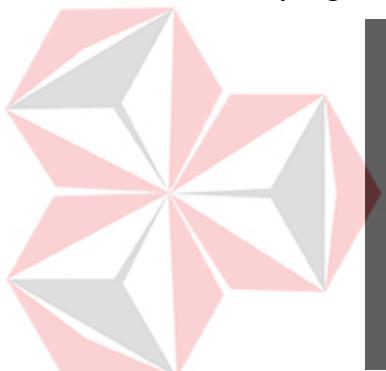
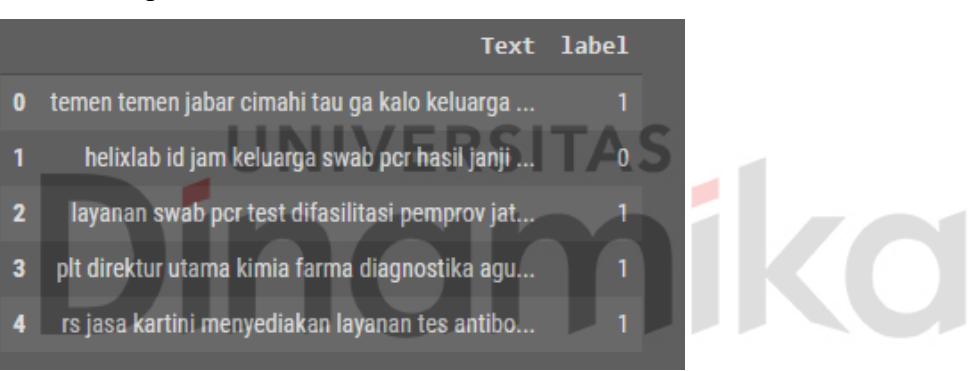
Gambar 4. 6 Hasil *Case Folding*

Kemudian dilanjutkan tahap *tokenizing* atau pemisahan kata pada python yang dapat dilihat pada Gambar L1.7 sampai Gambar L1.10. Sehingga hasil *tokenizing* akan ditampilkan 5 data teratas, dimana pada masing-masing data akan dipisahkan dengan tanda koma (,) untuk pemisahnya seperti pada Gambar berikut ini.

	Text	Label
0	[temen, temen, yang, di, jabar, khususnya, cim...	1
1	[halo, helixlab, id, sudah, jam, sejak, keluar...	0
2	[padahal, layanan, swab, pcr, test, yang, difa...	1
3	[plt, direktur, utama, kimia, farma, diagnosti...	1
4	[rs, jasa, kartini, menyediakan, layanan, tes,...	1

Gambar 4. 7 Hasil *Tokenizing*

Data pada kolom “Text” hasil penghapusan koma akan disimpan pada variabel bernama “*tweet*”, serta “df_preprocessed.head()” pada Gambar L1.11 digunakan untuk menampilkan 5 baris data teratas, berikut pada Gambar 4.8, 5 baris teratas yang berhasil ditampilkan

	Text	label
0	temen temen jabar cimahi tau ga kalo keluarga ...	1
1	helixlab id jam keluarga swab pcr hasil janji ...	0
2	layanan swab pcr test difasilitasi pemprov jat...	1
3	plt direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini menyediakan layanan tes antibo...	1

Gambar 4. 8 Hasil Penghapusan Tanda Koma

Setelah tahap *tokenizing*, dilanjutkan dengan *stopword removal*, untuk *code* proses *stopword removal* dapat dilihat pada Gambar L1.12 sampai Gambar L1.14. Sehingga dihasilkan hasil *stopword removal* seperti Gambar berikut ini.

	Text	label
0	temen temen yang di jabar khususnya cimahi ada...	1
1	halo helixlab id sudah jam sejak keluarga saya...	0
2	padahal layanan swab pcr test yang difasilitas...	1
3	plt direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini menyediakan layanan tes antibo...	1

Gambar 4. 9 Hasil *Stopword Removal*

Setelahnya *stopword removal* dilanjutkan proses terakhir pada tahap *text pre-processing* yaitu *stemming*. Tahap stemming pada python masih menggunakan *library* yang sama di tahap sebelumnya yaitu Sastrawi yang dapat dilihat pada Gambar L1.15. Kemudian `df_preprocessed.head()` pada baris terakhir *code* digunakan untuk menampilkan 5 baris pertama data set hasil *stemming*, sehingga dihasilkan hasil *stemming* seperti Gambar 4.10 berikut ini.

	Text	label
0	temen temen jabar cimahi tau ga kalo keluarga ...	1
1	helixlab id jam keluarga swab pcr hasil janji ...	0
2	layan swab pcr test fasilitas pemprov jatim tr...	1
3	plt direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini sedia layan tes antibodi sars ...	1

Gambar 4. 10 Hasil *Stemming*

4.5 Data *Training* dan Data *Testing*

Data hasil *crawling* akan dipecah atau dibagi menjadi data *training* dan juga data *testing* dengan perbandingan 80:20. Data *tweet* hasil *crawling* sejumlah 103 data yang akan dibagi menjadi data *training* dan *testing*. Data *training* akan digunakan untuk menyesuaikan model dengan kelas positif dan juga negatif. Data *training* yang digunakan jumlahnya lebih banyak yaitu sebanyak 80 % dari total keseluruhan data hasil *crawling*. Kemudian data *testing* akan digunakan untuk mengevaluasi kesesuaian model pada data *training* dengan perbandingan yaitu 20%.

Sehingga dengan pembagian tersebut dapat diketahui data *training* sejumlah 82 data *tweet* yang dapat dilihat pada Gambar 4.6, dan data *testing* sejumlah 21 data *tweet* dapat dilihat pada Gambar 4.7.

(83	tengah musim pandemi kait layan bantu maksimal...
72	situasi covid yg kait layan swab pcr audit kai...
65	desember gw positif swab pcr kabupaten rsud it...
92	bersih laboratorium tinjau lab sedia layan swa...
76	standar layan swab pcr tuh ga sih beda beda gi...
	...
88	cakup layan swab pcr puskesmas gak rata harap ...
3	plt direktur utama kimia farma diagnostika agu...
98	teman positif covid mg rawat rs kluar dr rs sw...
5	tri jabroni calon tumpang wajib salah surat te...
62	berita tv testing acak tempat umum cari otg la...
	Name: Text, Length: 82, dtype: object,

Gambar 4. 11 Data *Training*

84	yg tracing layan swab pcr waspada pasti alat b...
9	tri jabroni calon tumpang wajib salah surat te...
58	sebenarnya gw ga sih duit jg duit gw swab pcr ...
40	carolineputri mba bgt dr covid layan puskesmas...
8	bijak puskesmas beda kait layan swab pcr konta...
75	tengah naik kopet moga gaada korupsi sampe ala...
23	toniheru daftar apel pagi koordinasi mikir bag...
31	daily activities tugas puskesmas tahun tracing...
91	organisasi sehat dunia who umum pandemi tes sw...
25	layan swab pcr pintu suramadu arah sby hilang ...
47	dear pemkot depok layan swab pcr kontak dgn yg...
26	convomf rapid test murah kalo swab pcr an swab...
50	vaksin gratis warga indonesia kecuali layan gr...
19	buka kerjasama mitra sampling point swab pcr t...
53	upaya tingkat layan periksa swab pcr rs kartik...
101	suap nasi ganti swab pcr layan swab pcr sii yg...
37	negatif covid bukti hasil swab pcr negatif sur...
70	who anjur prosedur layan swab pcr but why not ...
28	si layan swab test pcr dateng jam nunggu dafta...
64	covid tingkat obat antivirus langka oximeter l...
42	moonareas swab tuh gak pake bpjs tau klinik co...

Gambar 4. 12 Data Testing

4.6 Pembobotan TF-IDF

Pembobotan TF-IDF akan dilakukan setelah data set melalui proses *text pre-processing*. Proses pembobotan TF-IDF pada python akan menggunakan *library* TfidfVectorizer, yang mana pembobotan akan dilakukan pada setiap kata data *training* dan data *testing*, dengan perbandingan keduanya 80:20. *Term Frequency* (TF) adalah Frekuensi kemunculan kata dalam sebuah dokumen yang menunjukkan seberapa penting kata tersebut dalam dokumen.

Kemudian *Inverse Document frequency* (IDF) atau bisa dikatakan frekuensi dokumen yang mengandung kata tersebut. Hasil dari pembobotan TF-IDF berupa data numerik yang kemudian akan digunakan pada proses perhitungan klasifikasi dengan SVM. Pembobotan TF-IDF dengan python memanfatkan *library* sklearn dengan fungsi TfidfVectorizer. Untuk *code* python pembobotan TF-IDF dapat dilihat pada Lampiran 3.Untuk hasil proses pembobotan TF-IDF dapat dilihat pada Gambar 4.13 berikut ini.

(0, 615)	0.16280453689448574
(0, 548)	0.2962335796606734
(0, 369)	0.2962335796606734
(0, 51)	0.2962335796606734
(0, 355)	0.07024146057030266
(0, 285)	0.27650751801961093
(0, 448)	0.3240359020218706
(0, 408)	0.3240359020218706
(0, 608)	0.3240359020218706
(1, 35)	0.4183785100329058
(1, 700)	0.22891262388485603
(1, 131)	0.19476887406652232
(1, 567)	0.4183785100329058

Gambar 4. 13 Hasil Pembobotan TF-IDF

Keterangan warna :

- | | |
|--------|--|
| Hijau | = <i>Index</i> kalimat ke-n dari data set |
| Kuning | = <i>Index</i> dari token yang terdapat pada kalimat tersebut. |
| Merah | = Bobot (hasil kalkulasi) |

4.7 Klasifikasi Dengan *Algoritma Support Vector Machine*

Setelah data melalui proses pelabelan kemudian akan dilakukan *text pre-processing*, yang mana pada tahap ini dilakukan pengolahan data. Data set kemudian akan diproses untuk klasifikasi SVM menggunakan *library* sk-learn. *Code* lengkap proses klasifikasi SVM dapat dilihat pada Lampiran 4.

Data *training* yang akan digunakan dalam perhitungan klasifikasi SVM berjumlah 82 data yang dibagi dalam 2 kelas yaitu positif (1) dan negatif (0). Hasil *training* pada algoritma SVM kemudian diuji dengan menggunakan data uji, sehingga didapatkan nilai akurasi dalam memprediksi pada data baru. Proses inilah yang disebut sebagai *machine learning*. Data *testing* yang akan diklasifikasi terlebih dahulu dilakukan *text pre-processing*, kemudian juga pembobotan TF-IDF. Hasil uji data *testing* pada klasifikasi SVM dapat dilihat pada Gambar 4.14 berikut ini.

	Text	label
0	temen temen jabar cimahi tau ga kalo keluarga ...	1
1	helixlab id jam keluarga swab pcr hasil janji ...	0
2	layan swab pcr test fasilitas pemprov jatim tr...	1
3	plt direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini sedia layan tes antibodi sars ...	1
...
98	teman positif covid mg rawat rs kluar dr rs sw...	0
99	layan tes covid genose swab pcr layan tsb swab...	0
100	nakes rodii igd gak sepi ruang isolasi tambah t...	0
101	suap nasi ganti swab pcr layan swab pcr sii yg...	0
102	day ppkm granostic dapat layan swab pcr udah d...	0

103 rows x 2 columns

Gambar 4. 14 Hasil Uji SVM Pada Data *Testing*

4.8 Evaluasi dan Validasi

Tahap evaluasi dilakukan untuk mengetahui hasil akurasi, waktu pemrosesan, dan kinerja dari algoritma *Support Vector Machine*. Kemudian validasi dilakukan oleh peneliti untuk mengetahui hasil akurasi dari algoritma SVM. Dalam hal ini *cross validation* dilakukan dengan *10-fold cross validation*, Dalam 10 fold *cross validation*, data dibagi menjadi 10 fold berukuran kira-kira sama, sehingga kita memiliki 10 subset data untuk mengevaluasi kinerja model atau algoritma. Untuk masing-masing dari 10 subset data tersebut, *cross validation* akan menggunakan 9 fold untuk pelatihan dan 1 fold untuk pengujian. Proses *10-fold cross validation* dengan python dapat dilihat pada Lampiran 5.1. Hasil *10-fold cross validation* akan ditampilkan dalam bentuk array seperti pada Gambar 4.15 berikut ini.

```
[ '66.7%', '55.6%', '37.5%', '25.0%', '62.5%', '50.0%', '50.0%', '62.5%', '75.0%', '75.0%' ]
```

Gambar 4. 15 Hasil *10-fold Cross Validation*

Kemudian untuk memperoleh hasil rata-rata dari *cross validation*, dalam hal ini digunakan fungsi “mean”, seperti yang terlihat di *code* pada Lampiran 5.1 pada Gambar 4.16. Sehingga didapatkan nilai rata-rata seperti pada gambar berikut ini.

56.0%

Gambar 4. 16 Rata-rata *10-fold Cross Validation*

Selanjutnya akan dilakukan prediksi pada data *testing* dengan python, dimana pada tahap ini akan menggunakan *library* sys dan numpy, untuk *code* proses

dari tahap prediksi data *testing* dapat dilihat pada Lampiran 5.2. Berikut Gambar merupakan hasil dari perintah print pada variabel “predict”. Dimana pada baris pertama merupakan susunan array dari hasil prediksi data *testing*, dan pada baris kedua merupakan length atau banyaknya data *testing* seperti pada Gambar 4.17.

```
[0 1 0 1 0 1 0 1 1 1 1 1 0 1 1 0 1 0 0 1 0]
21
```

Gambar 4. 17 Hasil Prediksi Data *Testing*

Selanjutnya akan dilanjutkan untuk tahap evaluasi yaitu dengan *confusion matrix*. *Code* proses evaluasi *confusion matrix* dapat dilihat pada Lampiran 5.3. Sehingga akan tampil hasil perintah print, seperti pada Gambar 4.18 berikut ini.

```
f1 score hasil prediksi      : 0.7826086956521738
accuracy score hasil prediksi : 0.7619047619047619
precision score hasil prediksi : 0.75
recall score hasil prediksi   : 0.8181818181818182
```

Gambar 4. 18 f1, Accuracy, Precision, & Recal Score

Kemudian dalam evaluasi terhadap algoritma *Support Vector Machine* dengan *confusion matrix* didapatkan *F1 score*, *accuracy*, *precision*, dan *recall*. Pada Gambar 4.14 F1 Score yang diperoleh 0.78, dimana hasil ini didapatkan dari perbandingan rata-rata *precision* dan *recall* yang diabotkan dengan rumus : $2 \times (recall \times precision) / (recall + precision)$, sehingga dihasilkan 0.78 atau 78%. *Accuracy score* mendapatkan nilai sebesar 0.76, *accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data, yang dihitung dengan rumus : $(TP + TN) / (TP + FP + FN + TN)$, sehingga menghasilkan 0.76 atau 76%. Kemudian pada *precision score* didapatkan nilai sebesar 0.75, yang merupakan rasio prediksi benar positif dibandingkan dengan hasil prediksi positif, yang dihitung dengan rumus : $(TP) / (TP + FP)$, sehingga menghasilkan 0.75 atau 75% nilai dari *precision*. Serta, pada *recall score* diperoleh nilai sebesar 0.82, yang merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif, yang dihitung dengan rumus : $(TP) / (TP + FN)$, sehingga menghasilkan nilai 0.82 atau 82% *recall score*.

Selanjutnya proses prediksi dengan *confusion matrix* akan menghasilkan hasil prediksi *true negative*, *false positive*, *false negative*, dan *true positive*, untuk *code*

lengkap fungsi prediksi dapat dilihat pada Lampiran 5.4. Hasil dari prediksi dapat dilihat pada Gambar 4.19 berikut ini.

Hasil True Negative	:	7
Hasil False Positive	:	3
Hasil False Negative	:	2
Hasil True Positive	:	9

Gambar 4. 19 Hasil Prediksi *Confusion Matrix*

Pada gambar 4.14 didapatkan nilai *True Negatif* (TN) = 7, yang menunjukan bahwa data negatif yang diprediksi benar terdapat 7 data. Kemudian pada hasil *False Positive* (FP) = 3 yang menunjukan, data negatif namun diprediksi sebagai data positif yaitu berjumlah 3 data. Lalu hasil *False Negatif* (FN) = 2 menunjukan, dimana data positif yang diprediksi sebagai data negatif dengan jumlah 2 data. Serta hasil *True Positive* (TN) = 9 menunjukan, bahwa terdapat 9 data positif yang diprediksi benar.

Dari hasil evaluasi dan validasi pada data yang semula dipisahkan antara kolom “Text” dan “Label”, supaya dapat digunakan untuk proses visualisasi maka kedua kolom akan digabungkan kembali pada Lampiran 5.5.

4.9 Visualisasi

Pada tahap ini data yang telah diproses hingga validasi akan divisualisasikan dengan penyajian yaitu dalam dua bentuk penyajian *Wordcloud* dan *Pie Chart*. Data set yang akan di sajikan dalam visualisasi, label pada dat set terlebih dahulu akan diubah menjadi numerik atau dengan 1 dan 0. Untuk *code* merubah label dapat dilihat pada Gambar L6.1. Sehingga label pada data set akan berubah menjadi numerik, seperti pada Gambar 4.22 berikut ini.

	Text	Label
0	temen temen jabar cimahi tau kalo keluarga tin...	1
1	helixlab id jam keluarga swab pcr hasil janji ...	0
2	layan swab pcr test fasilitas pemprov jatim tr...	1
3	plt direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini sedia layan tes antibodi sars ...	1

Gambar 4. 20 Hasil Perubahan Label Menjadi Numerik

4.9.1 Wordcloud

Sebelum proses visualisasi *wordcloud*, karena *wordcloud* merupakan kumpulan kata, terlebih dahulu *download* dan *import* stopword dari *library* *nltk* seperti pada Gambar L6.2. Kemudian untuk menghindari kemunculan kata redundant atau sama dari kedua *wordcloud* sehingga terlebih dahulu untuk memasukkan *word list positive* dan *negative*, supaya kata yang muncul pada *wordcloud* sesuai dengan *word list* masing-masing. *Word list* yang tersedia dalam format “.txt” terlebih dahulu untuk di *upload* ke dalam *drive*, untuk *code import* wordlist dapat dilihat pada Gambar L6.3 dan Gambar L6.4,

Setelah itu untuk mengolah data set menjadi *wordcloud* diperlukan sebuah *library* yaitu *wordcloud*, serta *library* *matplotlib* yang di beri alias menjadi plt untuk menampilkan hasilnya dapat dilihat pada Gambar L6.5 Serta untuk menampilkan *wordcloud* negatif berdasarkan kolom “Label” pada data set *code* dapat dilihat pada Gambar L6.6. dan Gambar L6.7. Hasil visualisasi *wordcloud* dapat dilihat pada Gambar 4.23 untuk sentimen negatif dan Gambar 4.24 untuk sentimen positif.



Gambar 4. 22 Wordcloud
Sentimen Negatif

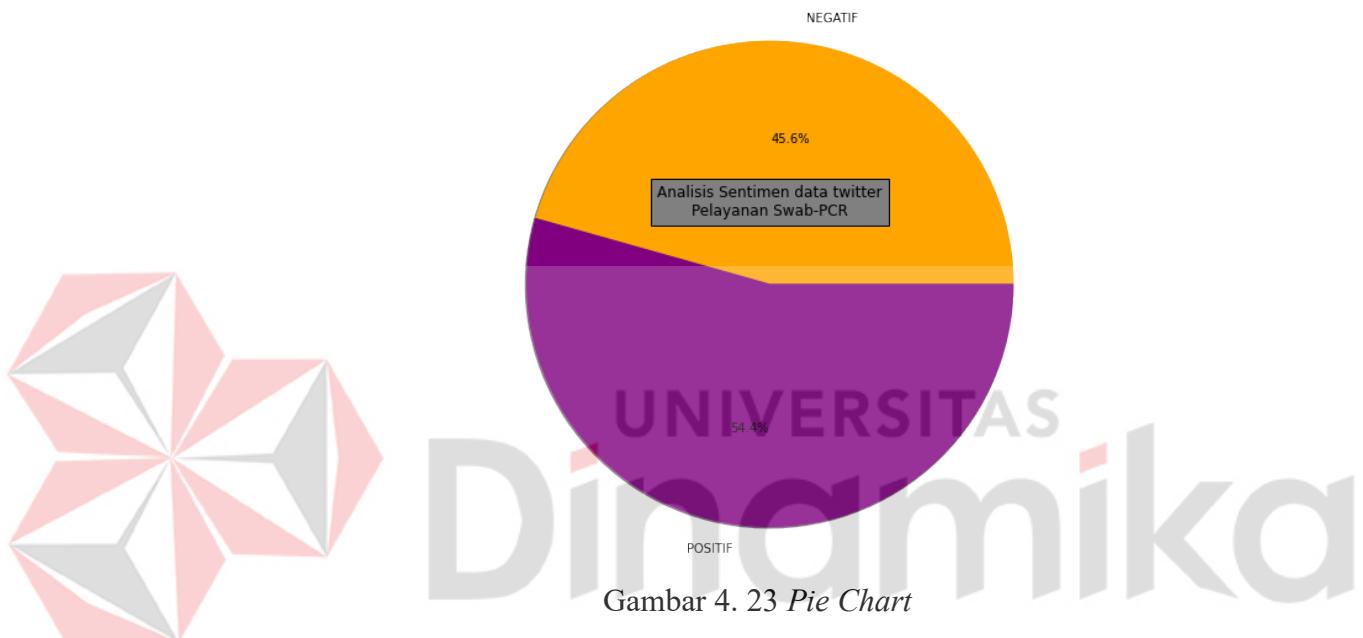


Gambar 4. 21 *Wordcloud*
Sentimen Positif

Pada Gambar 4.23 menunjukkan *wordcloud* sentimen negatif, dengan kata yang sering muncul ialah kecewa, mahal, tolak, dan antri. Dimana kata-kata tersebut berdasarkan *word list* negatif yang menunjukkan bahwa kata tersebut sering digunakan pengguna Twitter untuk membuat *tweet* yang telah berlabel sebagai data negatif. Pada gambar 4.23 menunjukkan *wordcloud* sentimen positif, dengan empat kata yang sering muncul ialah gratis, mandiri, sehat, dan positif. Dimana kata-kata tersebut berdasarkan *word list* positif yang menunjukkan bahwa kata tersebut sering digunakan pengguna Twitter untuk membuat *tweet* yang telah berlabel positif.

4.9.2 Pie Chart

Pada python visualisasi *pie chart* akan menggunakan *library* matplotlib yang dapat dilihat pada Lampiran 6.2. Persentase sentimen pada data *tweet* pelayanan *swab-PCR* sebesar 54,4% untuk positif dan 45,6 % untuk negatif. Berdasarkan “*type_count*” dimana fungsi *code* tersebut menghitung label positif = 56 data dan negatif = 47 data yang dibuktikan dengan perhitungan 54,4 % dari 103 = 56,2 jika dibulatkan = 56, begitu pula 45,6% dari 103 = 46,9 atau jika dibulatkan = 47. Hasil visualisasi *pie chart* dapat dilihat pada Gambar 4.17 berikut ini.



4.9.3 Rekomendasi Hasil Visualisasi

Jika diperhatikan dalam *wordcloud* negatif, empat kata yang sering muncul ialah kecewa, mahal, tolak, dan antri. Berikut data *tweet* yang mengandung keempat unsur kata tersebut pada Tabel 4.1 berikut ini.

Tabel 4. 1 Data Tweet Dengan Unsur Kata Negatif

Data ke-	Tweet	Unsur kata
49	Kecewa sekali dengan tahu pelayanan yang di faskes kesehatan d kota tercinta tangsel dimana fas SWAB PCR ya memang gratis di puskesmas tapi ya masa 2 minggu keluarnya? Udh keburu kemana mana itu virus kalo yang d tes beneran positif.	Kecewa
95	<i>Swab-PCR</i> sampai saat ini yang efektif, apakah pemerintah tidak berinisiatif untuk memberi subsidi terkait biaya pelayanan <i>swab-PCR</i> yang dirasa sangat mahal ini kepada masyarakat yang kurang mampu.	Mahal
83	di jawa timur Sejak akhir Desember 2020 hingga awal Januari 2021, LaporCovid-19 menerima total 23 laporan	Tolak

Data ke-	<i>Tweet</i>	Unsur kata
	kasus pasien yang ditolak mendapatkan pelayanan <i>swab pcr</i> krn RS penuh, pasien yang meninggal di perjalanan, serta di rumah karena ditolak RS. Pelayanan lagi genting.	
31	Kemarin nemenin suami PCR <i>drive-thru</i> di PremierBintaro antrian tidakberes banget deh. Masa antri sampe 2 jam. tidakdiarahkan pakai jalur yang mana. Jadi bisa aja yang datang duluan malah dapat pelayanan duluan. Baru di <i>swab</i> jam hampir jam 2 kurang bbrp menit.	Antri

Lalu dalam *wordcloud* positif, empat kata yang sering muncul ialah gratis, mandiri, sehat, dan positif. Berikut data *tweet* yang mengandung keempat unsur kata tersebut pada Tabel 4.2 berikut ini.

Tabel 4. 2 Data Tweet Dengan Unsur Kata Positif

Data ke-	<i>Tweet</i>	Unsur kata
50	Pengalaman langsung temen yang positif covid. Doi gratis biaya perawatan di rumah sakit swasta dgn pelayanan yang sangat baik. Temen2 kerja nya yang kontak langsung dgn dia, sudah tes <i>swab</i> PCR langsung di puskesmas gratis.	Gratis
33	Negatif covid-19 yang dibuktikan dengan hasil <i>swab</i> PCR negatif atau surat keterangan telah selesai isolasi mandiri dari fasilitas pelayanan kesehatan.	Mandiri
53	Pemeriksaan <i>swab pcr</i> covid di rumah sakit memang sudah jadi standar pelayanan. Tujuannya jelas. Untuk melindungi pasien dan lingkungan pasien termasuk tenaga kesehatan dan pasien lainnya, dan untuk memberikan terapi yang sesuai	Sehat
62	pelayanan <i>swab</i> dan pemeriksaan PCR yang dilakukan oleh Dinkes tidak dipungut biaya alias gratis. Namun <i>swab</i> tersebut ditujukan untuk kontak erat hasil tracing kasus konfirmasi positif yang dilakukan tim tracer surveilans puskesmas dan tim detektif covid kelurahanRW Siaga Covid-19	Positif

Berdasarkan persentase pada *pie chart* didapatkan sentimen negatif sebesar 45,6%, dimana pada *wordcloud* berdasarkan data *tweet* didapati ungkapan diantaranya, kekecewaan atas pelayanan *swab-pcr*, tarif tes yang mahal bagi masyarakat yang kurang mampu sehingga diharapkan subsidi, penolakan melayani pasien yang ingin melakukan tes *swab-pcr* karena penuh, sehingga dibutuhkan tenaga medis dan tumpat tambahan untuk menampung pasien yang lebih banyak, serta pasien harus antri hingga 2 jam dikarenakan kurangnya tenaga medis dan infrastruktur pelayanan.

Didapatkan 54,4% dari *pie chart* pada sentimen positif, dimana pada *wordcloud* berdasarkan data *tweet* didapati ungkapan melalui *tweet* diantaranya ialah, pelayanan kesehatan gratis untuk pasien yang positif covid-19 termasuk pelayanan *swab-pcr*, negatif covid-19 dapat dibuktikan dengan surat isolasi mandiri dan hasil *swab* yang negatif tentunya, tenaga kesehatan juga perlu memperoleh pelayanan *swab-pcr*, serta pelayanan *swab-pcr* di gratiskan oleh Dinkes untuk masyarakat yang ter konfirmasi positif.

4.10 Kesimpulan dan Saran

Berdasarkan proses analisis yang telah dilakukan diatas, yang dimulai dari tahap analisis, *crawling* data, pelabelan, *text pre-processing*, pembagian data, pembobotan TF-IDF, tahap klasifikasi SVM, evaluasi dan validasi, serta yang terakhir yaitu visualisasi. Dapat disimpulkan dari hasil proses klasifikasi maka didapatkan skor validasi dengan *cross validation* sebesar 66% yang menandakan nilai ketepatan model terhadap data *training*. Kemudian didapatkan skor pengujian pada nilai *f1 score* sebesar 78%, *accuracy* sebesar 76%, *precision* sebesar 75%, dan *recall* sebesar 81%. Selanjutnya pada *pie chart* divisualisasikan dengan perolehan persentase sentimen positif sebesar 54,4% dan persentase sentimen negatif sebesar 45,6%. Serta saran untuk penelitian serupa yang diharapkan dapat lebih baik kedepannya. Diharapkan untuk melakukan *crawling* pada waktu yang tepat dengan kondisi data yang dibutuhkan. Menggunakan *library stopword removal* dan *stemming* yang lebih baik supaya menghasilkan *output* yang lebih akurat. Dan diharapkan pada penelitian selanjutnya proses pelabelan data dilakukan secara otomatis, terlebih jika data set yang digunakan dalam jumlah besar.

BAB V

PENUTUP

Pada bab ini penulis akan membahas kesimpulan yang diperoleh dari hasil analisis dan pengujian yang telah dilakukan terkait tugas akhir ini, serta saran yang berguna untuk pengembangan penelitian serupa yang lebih baik kedepannya .

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan sentimen terhadap opini publik pada *tweet* dari Twitter, dengan memanfaatkan algoritma *Support Vector Machine*, *keyword* pelayanan *swab-PCR*, serta evaluasi dan validasi dengan *cross validation* dan *confusion matrix* maka didapatkan beberapa kesimpulan ialah sebagai berikut:

1. Score validasi didapatkan dari *cross validation* yaitu pada *fold* 1 sebesar 0.66, *fold* 2 sebesar 0.78, *fold* 3 sebesar 0.37, *fold* 4 sebesar 0.25, *fold* 5 sebesar 0.75, *fold* 6 sebesar 0.62, *fold* 7 sebesar 0.625, *fold* 8 sebesar 0.75, *fold* 9 sebesar 1, dan *fold* 10 sebesar 0.75, serat didapatkan rata-rata sebesar 0.66, dimana angka tersebut menunjukkan bahwa model klasifikasi masih belum sempurna (100%) yang disebabkan oleh bahasa gaul terdapat dalam data tweet yang tidak dapat di normalisasi sehingga tidak dikenali oleh kamus/sistem.
2. Didapatkan nilai pengujian pada nilai f1 *score* sebesar 78% dimana hasil tersebut didapatkan dari *precision* dan *recall*, serta jika hasil f1 score baik megindikasikan bahwa model klasifikasi yang dimiliki baik, *accuracy* sebesar 76% yang menunjukkan berapa persen data *tweet* yang diprediksi negatif dan positif dari keseluruhan data set, *precision* sebesar 75% yang menunjukkan berapa persen data *tweet* negatif yang diprediksi negatif, dan *recall* sebesar 81% yang menunjukkan berapa persen data *tweet* yang diprediksi negatif dibandingkan keseluruhan data *tweet* yang sebenarnya negatif.
3. Berdasarkan visualisasi *wordcloud* didapatkan empat kata positif yang sering muncul ialah “gratis”, “mandiri”, “sehat”, dan “positif” didapati ungkapan melalui *tweet* diantaranya ialah, pelayanan kesehatan gratis untuk pasien yang positif covid-19 termasuk pelayanan *swab-pcr*, negatif covid-19 dapat dibuktikan dengan surat isolasi mandiri dan hasil *swab* yang negatif tentunya, tenaga kesehatan juga perlu memperoleh pelayanan *swab-pcr*, serta pelayanan

swab-pcr di gratiskan oleh Dinkes untuk masyarakat yang ter konfirmasi positif. Serta empat kata negatif yang paling sering muncul ialah “kecewa”, “mahal”, “tolak”, dan “antri” didapati ungkapan diantaranya, kekecewaan atas pelayanan *swab-pcr*, tarif tes yang mahal bagi masyarakat yang kurang mampu sehingga diharapkan subsidi, penolakan melayani pasien yang ingin melakukan tes *swab-pcr* karena penuh, sehingga dibutuhkan tenaga medis dan tumpat tambahan untuk menampung pasien yang lebih banyak, serta pasien harus antri hingga 2 jam dikarenakan kurangnya tenaga medis dan infrastruktur pelayanan.

4. Dari total 103 data tweet didapatkan persentase sentimen positif sebesar 54,4% atau sejumlah 56 *tweet* positif, dan persentase sentimen negatif sebesar 45,6% atau sejumlah 44 *tweet* negatif yang divisualisasikan melalui *pie chart* berdasarkan data set dengan *keyword* pelayanan *swab pcr*.

5.2 Saran

Adapun saran yang dapat digunakan untuk pengembangan penelitian selanjutnya antara lain:

1. Diharapkan pada penelitian selanjutnya data set yang akan digunakan dapat dilakukan pelabelan otomatis.
2. Pada penelitian ini, data set yang digunakan sejumlah 103, diharapkan pada penelitian berikutnya data set yang digunakan lebih banyak, supaya tingkat akurasi menjadi lebih tinggi.
3. Pada penelitian selanjutnya diharapkan dapat menggunakan pendekatan machine learning yang lain untuk membandingkan performa algoritma SVM.
4. Diharapkan pada penelitian selanjutnya untuk melakukan *crawling* pada waktu yang tepat dengan kondisi data yang dibutuhkan.
5. *Library* Sastrawi untuk *text pre-processing* pada tahap *stopword removal* dan *stemming* kurang spesifik dalam menghilangkan kata tidak penting dan mengubah kata dasar diharapkan pada penelitian selanjutnya *library* untuk *stopword* dan *stemming* untuk menggunakan *library* lain yang lebih akurat.

DAFTAR PUSTAKA

- Anjasmoros, M. T., Istiadi, & Marisa, F. (2020). ANALISIS SENTIMEN APLIKASI GO-JEK MENGGUNAKAN METODE SVM DAN NBC (STUDI KASUS: KOMENTAR PADA PLAY STORE). *Conference on Innovation and Application of Science and Technology (CIASTECH 2020)*.
- dr. V. Fridawati, S. (2020). *Apa Itu Rapid Test, Swab, dan PCR, apakah tahu Perbedaannya?* Retrieved from PRIMAYA HOSPITAL: <https://primayahospital.com/>
- Flawcett, T. (2006). In An Introduction to ROC Analysis. *Pattern Recognition Letters*, 861-874.
- Francis, L., & Flynn, M. (n.d.). *Text Mining Handbook*, Casualty Actuarial Society E-Forum. Spring: p.l.
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *JURNAL SAINS DAN SENI ITS Vol. 6, No. 2 (2017)*, 2337-3520.
- KawalCOVID19.id, T. a. (2020, April 27). *Rapid test atau swab test: Apa bedanya? Mana yang lebih baik?* Retrieved from KAWAL COVID19: <https://kawalcovid19.id/>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- Loelianto, I., Sofyan, M., & Angriani, H. (2020). IMPLEMENTASI TEORI NAÏVE BAYES DALAM KLASIFIKASI CALON MAHASISWA BARU STMIK KHARISMA MAKASSAR. *SINTECH JOURNAL*, 113.
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Makarim, d. F. (2020, October 20). *PCR Test dan Swab Antigen Tidak Sama, Ini Penjelasannya*. Retrieved from halodoc: <https://www.halodoc.com/>

- Neneng, N., Adi, K., & Isnanto, R. (2016). Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM). *Jurnal Sistem Informasi Bisnis*.
- Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: DEEPUBLISH (Grup Penerbitan CV BUDI UTAMA).
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *e-Proceeding Of Engineering : Vol.2, No.1 April 2015 | Page 1177*, 1177.
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2790.
- Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Feature. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- Supriyatna, A. (2018). Metode Extreme Programming Pada Pembangunan Web Aplikasi Seleksi Peserta Pelatihan Kerja. *Jurnal Teknik Informatika*, 11(1), 1-18.
- Taufik, & Pamungkas, S. (2018). ANALISIS SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) .
- Tiara, Sabariah, M., & Effendy, V. (2015). Sentiment analysis on Twitter using the combination of lexicon vector machine for assessing the performance of a television program. 3rd International Conference on Information and Communication Technology. *ICoICT*, 386-390.
- Utomo, M. S. (2013). Implementasi Stemmer Tala pada Aplikasi Berbasis Web. *Jurnal Teknologi Informasi DINAMIK*.

Yuhefizar, M., & Hidayat, R. (2009). *Cara Mudah Membangun Website Interaktif Menggunakan Content Management System*. Jakarta: PT. Elex Media Komputindo.

Zhang, L., Ghosh, R., Dekhil, Hsu, M., & Liu, B. (2011). Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. 89(HPL-2011-89,).

