

SVM

by Amirul Mukminin

Submission date: 14-Jul-2021 04:15PM (UTC+0700)

Submission ID: 1619487583

File name: jurnal-17410100189.docx (776K)

Word count: 4485

Character count: 28416

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

Amirul Mukminin 1) Dewiyani Sunarto 2) Vivine Nurcahyawati 3)

Program Studi/Jurusan Sistem Informasi
Universitas Dinamika

Jl. Raya Kedung Baruk 98 Surabaya, 60298

Email : 1) dev.mukminin@gmail.com, 2) dewiyani@Dinamika.ac.id, 3) vivine@Dinamika.ac.id

Abstract: One of the Covid-19 tests advocated through the World Health Organization (WHO) is Swab-PCR which is considered to have high accuracy compared to other test methods. Various responses related to this were conveyed by the public through social media, one of which was Twitter. This study uses data from Twitter regarding the opinion of the swab-PCR test service in Indonesia, Then the analysis is carried out to produce classifications and predictions that are expected to help the government in improving PCR-swab services. This study will conduct a sentiment analysis by classifying public opinion from Twitter into positive opinions and negative opinions. The classification process begins by assigning weights using the Term Frequency Inverse Document Frequency (TF-IDF) algorithmic rule and also the Support Vector Machine utilized classification process, as well as a validation with an average score of 0.66 which is then tested with the acquisition of a score, accuracy of 76%, precision of 75%, and recall of 81%. The percentage of positive sentiment is 54.4% and the percentage of negative sentiment is 45.6% which is visualized through a pie chart based on a data set with the keyword "PCR swab service".

Keywords: Sentimen analysis, SWAB-PCR services, Support Vector Machine, SVM

2
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* 2 *Real Time - Polymerase Chain Reaction (PCR)* lebih akurat. Metode ini (*swab-PCR*) 2 merekomendasikan *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 respon dan opini publik (masyarakat) terhadap pelayanan *swab-PCR*. Hal tersebut sangat perlu dilakukan, karena pemerintah dapat mempertimbangkan pada saat menyikapi respon atau opini publik.

Opini atau tanggapan publik bisa dipandang melalui banyak sekali media sosial. Media sosial seolah sebagai suatu keharusan memiliki oleh setiap orang pada *device* yang dimiliki. Menurut konten HootSuit dan We Are Social, dengan laporan yang judulnya "Digital 2021" di tahun 2021 Indonesia mempunyai 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 be 33 edia sosial, jumlah klien media sosial sama dengan 61,8% dari total populasi pada Januari 2021. Sejumlah 170 juta jiwa merupakan pengguna aktif media sosial. Media berbasis web yang sering digunakan 12 syarakat Indonesia salah satunya adalah 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.

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 respon 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 menggunakannya 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).

Respon atau opini masyarakat dari Twitter digunakan sebagai data penelitian, dan dibagi menjadi data *training* dan data *testing*. Data *training* sendiri digunakan untuk memperoleh parameter untuk pengklasifikasian dengan algoritma SVM sedangkan *testing* data digunakan untuk menyelesaikan siklus penilaian hasil karakterisasi (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 respon 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 sebagai bahan penilaian terhadap pelayanan *swab-PCR*. Berdasarkan permasalahan diatas, penelitian ini akan melakukan analisis sentimen untuk melakukan klasifikasi respon atau opini publik terhadap pelayanan *swab-PCR* menggunakan algoritma *Support Vector Machine*.

METODE

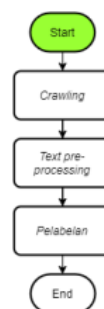
Tahapan yang digunakan dalam menyelesaikan penelitian ini terdiri dari tahapan awal, pengembangan dan akhir.

Tahap Awal



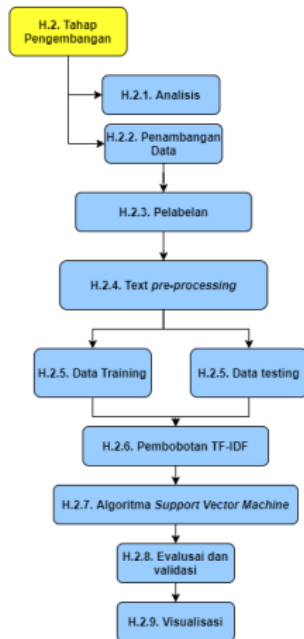
Gambar 1. Tahap Awal Penelitian

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 ini. Kemudian data hasil *crawling* ini dapat dikatakan masih belum bersih, dalam artian masih terdapat tanda baca, kata yang tidak penting, *link*, dan lain sebagainya. Oleh karena itu data hasil *crawling* terlebih dahulu akan dilakukan proses pembersihan yang disebut *text pre-processing*. Kemudian akan dilakukan dengan pelabelan, dengan dibagi pada 2 kelas yaitu kelas positif dan kelas negatif. Diagram alir teknik pengolahan data dapat dilihat pada gambar 2 berikut ini:



Gambar 2. Tahap Pengolahan Data

Tahap Pengembangan



Gambar 3. Tahap Pengembangan

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 dari identifikasi permasalahan, Analisis Kebutuhan Pengguna, Kebutuhan Fungsional, Kebutuhan Non-Fungsional, dan kebutuhan sistem. Tahap identifikasi masalah dilakukan dengan mempelajari studi literatur yang serupa dengan penelitian yang akan dikerjakan penulis, dengan melakukan observasi pada tweet pada Twitter berbahasa Indonesia dengan keyword “pelayanan swab per”, sehingga didapatkan informasi yang berkaitan dengan penelitian.

Analisis

Analisis Kebutuhan Pengguna dilakukan untuk mengetahui kebutuhan pengguna yang dibutuhkan dalam penelitian ini. berdasarkan hasil Identifikasi Masalah maka dapat dilakukan analisis kebutuhan pengguna untuk penelitian ini sebagai berikut:

Tabel 1. Kebutuhan Pengguna

Pengguna	Kebutuhan Pengguna
Peneliti	1. Survei dilakukan terhadap tweet pada Twitter berbahasa

Pengguna	Kebutuhan Pengguna
	Indonesia dengan keyword “pelayanan swab per”
	2. Melakukan pengelompokan komentar
	3. Menampilkan hasil dari pengelompokan komentar menurut kelompok positif dan negatif.
	4. Menampilkan kata yang sering muncul
	5. Menampilkan jumlah kata yang sering muncul

Tahap yang dilakukan selanjutnya adalah analisis kebutuhan fungsional yang 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.

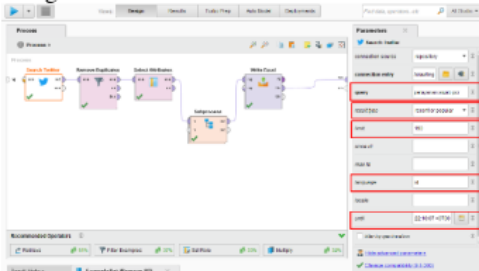
Analisis kebutuhan non-fungsional dilakukan untuk mengetahui prasyarat tidak berguna yang akan dimanfaatkan. Kemudian kebutuhan non-fungsional yang diperlukan adalah kebutuhan akses dan kebutuhan dalam menampilkan data dalam bentuk visualisasi dengan waktu yang efektif.

Dalam proses penambangan data digunakan tools 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 kurang lebih sejumlah 103 data, yaitu dengan melakukan sekali *crawling* dengan menggunakan RapidMiner.

Pengembangan Data

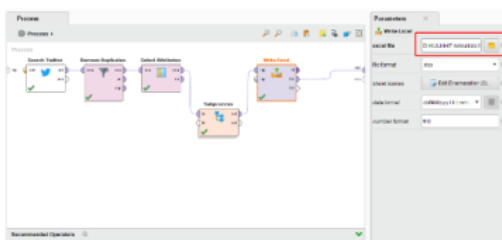
Penambangan data dilakukan dengan melakukan pencarian data (*crawling*) pada Twitter. Tahap *crawling* menggunakan tools RapidMiner, yaitu dengan menghubungkannya dengan Twitter menggunakan API Twitter. API diperoleh dengan mengakses Twitter developer yang kemudian dari Twitter developer tersebut akan memberikan access token. Access token sejumlah 6 digit yang didapatkan akan diinputkan pada RapidMiner. Pada RapidMiner waktu *crawling* dapat ditentukan hingga 7 hari kebelakang yang dapat di kustomisasi 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 *maximum* 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 mengisi kolom *language* dengan “id”, id menginisialisasi bahasa Indonesia.



Gambar 4. Parameter Crawling

Setelah parameter ditentukan, langkah berikutnya ialah memberikan operator *remove duplicate* untuk menghapus data yang sama dalam operator ini parameter yg 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 diisi *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 5.



Gambar 5. Menyimpan Data Crawling.

Setelah operator *crawling* tersusun, proses *crawling* dapat dijalankan dengan menekan *button run*. *Crawling* dilakukan 4 kali dalam rentang waktu 7 hari, dari hasil 4 kali *crawling* dapat dikalkulasikan data yang diperoleh kurang

lebih 103 data. Dari hasil *crawling* data disimpan dalam file Excel, untuk kemudian akan dilakukan pelabelan.

Pelabelan

Pelabelan atau penentuan kelas atribut diberikan sesuai dengan objektifitas peneliti, pelabelan dilakukan oleh sukarelawan dengan tujuan pelabelan pada setiap *tweet* bersifat objektif. Pembagian 25 kelas yaitu kelas positif untuk data *tweet* yang bersifat positif dan kelas negatif untuk data *tweet* yang bersifat negatif. Pelabelan penentuan positif dan negatif ditentukan berdasarkan acuan *load dictionary*.

Tabel 2 Contoh Kamus Positif dan Negatif

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

Proses pelabelan data dilakukan dengan melibatkan 2 sukarelawan yang masing-masing sukarelawan akan melabeli 50% data set.

Text Pre-Processing

Tahap *text pre-processing* yang dilakukan dalam tahap ini adalah melakukan ekstraksi data. Ekstraksi data bertujuan untuk mengolah hasil 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*, *stoword removal*, dan *stemming*. Proses *text pre-processing* dapat dilihat pada gambar berikut ini.



Gambar 6. Diagram Alur Text Pre-Processing

Dalam *text pre-processing* data set akan masuk kedalam 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.

Pembagian Data

Data *training* dan data *testing* merupakan data set telaah dibagi dengan perbandingan 80:20 yang akan digunakan dalam penelitian ini. Pembagian porsi, sejumlah 80% untuk *training* dan sejumlah 20 % untuk *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 terlabeli, kemudian dilakukan proses *text pre-processing* untuk dijadikan bahan inputan dari penggunaan algoritma *Support Vector Machine*. Pada bagian proses *testing*, merupakan data yang tidak terlabeli atau data yang tidak termasuk data *training*. Meskipun data *testing* tidak diperlukan pelabelan tetap harus melalui proses *text pre-processing*. Ketika data *ing* 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 gambar berikut ini:

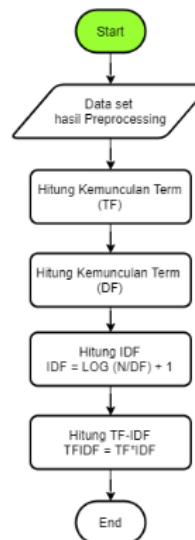


Gambar 7. Proses Pembagian Data

Pada gambar pembagian dilakukan dengan python dimana data training akan diambil sebanyak 80% dalam artian 82 data. Jika kondisi tersebut sudah terpenuhi maka sisa data akan dijadikan sebagai data *testing* yaitu sebanyak 20%.

Pembobotan TF-IDF

Pembobotan TF-IDF pada python akan menggunakan *library* *TfidfVectorizer* untuk mengubah data yang berupa kata/string menjadi numerik yang disertai dengan bobot di setiap data. Maka kemudian, demikian data numerik lah yang kemudian akan digunakan dalam proses berikutnya. Diagram alir proses pembobotan TF-IDF dapat dilihat pada gambar berikut:



Gambar 8. Pembobotan TF-IDF

Pada tahap penghitungan nilai TF-IDF diperlukan kata kunci sebagai parameter

perhitungan, kata kunci tersebut yaitu disesuaikan dengan judul penelitian yang dilakukan.

Tabel 3. Contoh Data Tweet

Kata Kunci	Pelayanan swab pcr
Data 1	sesuai soal layanan swab pcr
Data 2	layanan swab pcr kontak positif area antri mesti tunggu banget solusi tracing swab antigen

Klasifikasi Support Vector Machine

Pada tahap pengembangan penelitian, algoritma *Support Vector Machine (SVM)* akan diterapkan dalam tahap klasifikasi. Proses klasifikasi akan dilakukan dengan mempelajari data training yang telah dibentuk pada pembagian data *trainiing* dan data *testing*.

Evaluasi dan Validasi

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

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.

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil dan implementasi dari penelitian berdasarkan tahapan-tahapan yang sudah dilakukan sesuai dengan metode yang digunakan.

Analisis

Analisis dimulai dengan tahap penambangan data/*crawling*, *text pre-processing*,

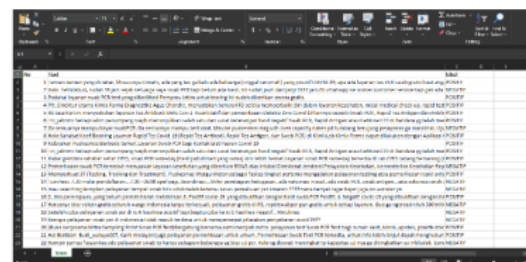
kemudian dilakukan pembagian data untuk mendapatkan data *training* dan *testing*, lalu perhitungan sentimen manual, pelabelan manual, pengklasifikasian dengan algoritma SVM, dan melakukan visualisasi berupa *wordcloud* dan *pie chart*. Penelitian ini dilakukan dengan *tools* RapidMiner, dan juga Google Colabs. 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.

Penambangan Data

Setelah dilakukanya 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).

Pelabelan

Setelah didaptkannya data hasil *text pre-processing* maka kemudian akan dilakukan pelabelan, pelabelan dilakukan pada data hasil *crawling*. Pelabelan dilakukan kurang lebih pada 103 data yang didapatkan dari hasil empat kali *crawling* pada media sosial Twitter. Setiap sekali *crawling* akan mendapatkan 25 sampai dengan 35 data *tweet* yang mengandung *keyword*. Pelabelan akan dilakukan oleh 2 sukarelawan dan masing masing sukarelawan akan melabeli 52 data, dan memberikan kelas pada masing-masing data yaitu kelas positif dan negatif. Untuk contoh beberapa data yang sudah dilakukan pelabelan oleh sukarelawan.



Gambar 9 Hasil Pelabelan.

Text Pre-Processing

Setelah melalui proses pelabelan dan sebelum proses *text pre-processing* label akan diganti POSITIF = 1 dan NEGATIF = 0 secara

otomatis melalui python, penggantian label dengan angka dilakukan untuk pembobotan TF-IDF data hasil penambangan (*crawling*) akan dilakukan *text pre-processing*. Pada tahap ini data hasil *crawling* yang formatnya tidak terstruktur akan diolah sehingga menghasilkan kata dasar dengan format yang sama, sehingga memudahkan proses berikutnya. *Text pre-processing* sendiri terdapat 5 tahapan yaitu *case folding*, *cleansing*, *tokenizing*, *stopword removal*, dan *stemming*.

Text pre-processing pada python menggunakan beberapa *library*, nltk digunakan untuk proses *tokenizing*, *library sastrawi* digunakan untuk proses *stopword removal* dan proses *stemming*. Data yang digunakan proses *text pre-processing* dapat dilihat 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
3	Pit Direktur Utama Kimia Farma Diagnostika Ag...	1
4	RS Jasa Kartini menyediakan layanan Tes Antibo...	1
5	tri jabroni Setiap calon penumpang wajib menu...	1

Gambar 10 Data *Text Pre-processing*

Dari data pada gambar diatas akan dilakukan proses pertama dari tahap *text pre-processing* yaitu proses *case folding*. Hasil dari tahap *case folding* dapat dilihat 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
3	pit direktur utama kimia farma diagnostika ag...	1
4	rs jasa kartini menyediakan layanan tes antibo...	1
5	tri jabroni setiap calon penumpang wajib menu...	1

Gambar 11 Hasil *Case Folding*

Setelah melewati tahap *case folding*, kemudian data akan dilanjutkan untuk proses *cleansing*. Hasil dari tahap *cleansing* dapat dilihat 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
3	Pit Direktur Utama Kimia Farma Diagnostika Ag...	1
4	RS Jasa Kartini menyediakan layanan Tes Antibo...	1
5	tri jabroni Setiap calon penumpang wajib menu...	1

Gambar 12 Hasil *Cleansing*

Proses *text pre-processing* berikutnya ialah *tokenizing*, dimana data yang berbentuk akan dipisahkan dengan koma (,) sehingga susunan kata. Hasil dari tahap *tokenizing* dapat dilihat 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	[pit, direktur, utama, kimia, farma, diagnosti...	1
4	[rs, jasa, kartini, menyediakan, layanan, tes,...	1
5	[tri, jabroni, setiap, calon, penumpang, wajib,...	1

Gambar 13 Hasil *Tokenizing*

Kemudian setelah proses *tokenizing*, proses selanjutnya ialah *stopword removal*, *library Sastrawi* akan digunakan dalam tahap *stopword removal*, dimana kata-kata yang tidak penting pada masing-masing data akan dihilangkan dengan mencocokkannya dengan daftar kata tidak penting (*stopword*). Kata tidak penting pada bahasa Indonesia contohnya seperti, dan, dengan, yang, di, ke, dan lain sebagainya. Hasil dari tahap *Stopword Removal* dapat dilihat pada gambar 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	layanan swab pcr test difasilitasi pemprov jat...	1
3	pit direktur utama kimia farma diagnostika agu...	1
4	rs jasa kartini menyediakan layanan tes antibo...	1
5	tri jabroni calon penumpang wajib salah surat ...	1

Gambar 14. Hasil *Stopword Removal*

Setelah proses *stopword removal*, terakhir pada proses preprocessing ialah *stemming*, pada python proses *stemming* akan menggunakan *library Sastrawi*, dimana setiap kata pada data akan diubah kedalam bentuk kata dasar, imbuhan pada setiap kata akan dihilangkan,

sehingga didapatkan kata dasar yang sesuai dengan kata-kata bahasa Indonesia. Hasil dari tahap *stemming* dapat dilihat pada gambar berikut.

	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
5	tri jabroni calon tumpang wajib salah surat te...	1

Gambar 15 Hasil tahap stemming

Data Training dan Data Testing

Data hasil *crawling* akan dipecah atau dibagi menjadi data *training* dan data *testing* untuk perbandingan pembagian 80:20. Data hasil *crawling* sejumlah 103 data yang akan dibagi. Data *training* akan digunakan untuk mencocokkan 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* yaitu pada Gambar 16, dan data *testing* sejumlah 21 data *tweet* dapat dilihat pada Gambar 17.

(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 gl...
...	
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 ota la...

Name: Text, Length: 82, dtype: object,

Gambar 16. 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	siapa 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 datang jam nunggu dafta...
64	covid tingkat obat antivirus langka oximeter l...
42	moonareas swab tuh gak pake bpjs tau klinik co...

Gambar 17 Data Testing

Pembobotan TF-IDF

Tahap ini dilakukan setelah data set dilakukan *pre-processing*. Proses pembobotan TF-IDF pada python akan menggunakan library *TfidfVectorizer*, yang mana pembobotan akan dilakukan pada setiap kata pada data *training* dan *testing*, dengan perbandingan kedua 80:20. *Term Frequency* (TF) merupakan 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 memanfaatkan library *TfidfVectorizer*. Untuk hasil pemberian bobot dengan TF-IDF dapat dilihat pada gambar berikut ini.

(0, 455)	0.0710932598956497
(0, 588)	0.0710932598956497
(0, 495)	0.20117681165615398
(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
(1, 455)	0.09179196491170448
(1, 588)	0.09179196491170448
(1, 355)	0.18138433076440297
(1, 285)	0.7140246045583776

Gambar 18 Hasil Pembobotan TF-IDF

Keterangan warna :

- Hijau = Term Frequency
- Kuning = Inverse Document Frequency
- Merah = Bobot

Klasifikasi Dengan Algoritma Support Vector Machine

Proses klasifikasi akan dilakukan dengan cara mempelajari data training yang telah dibentuk pada sub bab sebelumnya. Setelah data melalui proses pelabelan kemudian akan dilakukan *text-preprocessing*, yang mana pada tahap ini dilakukan pengolahan data. Data set kemudian akan diproses untuk klasifikasi SVM menggunakan library *sk-learn*. Data *training* yang akan digunakan dalam perhitungan

klasifikasi SVM berjumlah kurang lebih 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*. Hasil klasifikasi dengan *Support Vector Machine* dapat dilihat pada gambar berikut ini.

(0, 112)	0.14056123011345237
(0, 132)	0.2254773122735065
(0, 148)	0.281603789631076
(0, 149)	0.25676742774518185
(0, 213)	0.281603789631076
(0, 314)	0.06209890153917937
(0, 341)	0.25676742774518185
(0, 347)	0.281603789631076
(0, 382)	0.281603789631076
(0, 394)	0.25676742774518185
(0, 403)	0.06295561487939522
(0, 416)	0.281603789631076
(0, 485)	0.17718114093345624
(0, 496)	0.25676742774518185

Gambar 19 Hasil Klasifikasi SVM

Hasil klasifikasi pada gambar diatas akan digunakan untuk mengklasifikasi data testing kedalam kelas positif dan kelas negatif. Data testing yang akan diklasifikasi terlebih dahulu dilakukan *text pre-processing*, kemudian juga pembobotan *tf-idf*. Hasil uji data testing pada klasifikasi *Support Vector Machine* dapat dilihat pada gambar 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	pjt 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 rodi igd gak sepi ruang isolasi tambah t...	0
101	suap nasi ganti swab pcr layan swab pcr sii yg...	0
102	day ppkm granostic dapet layan swab pcr udah d...	0

Gambar 20 Hasil Uji Data Testing

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 validasi ini penulis menggunakan teknik *cross-validation*

dengan membagi data training menjadi beberapa bagian, dan akurasi akan diukur dengan menggunakan *confusion matrix*.

Tabel 4 *Cross Validation*

Fold	Cross Validation Score
1	0.66666667
2	0.77777778
3	0.375
4	0.25
5	0.75
6	0.625
7	0.625
8	0.75
9	1
10	0.75
Average	0.65694

Hasil dari perhitungan akurasi akan menghasilkan *true positive* yang didapat dari pelabelan dan prediksi dengan hasil *positive*, untuk *false positive* didapatkan dari pelabelan dengan hasil *negative* dan prediksi *positive*, *false negative* didapatkan dari hasil pelabelan *positive* dan prediksi *negative*, serta *true negative* didapatkan dari hasil pelabelan dan prediksi yang *negative*.

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

Gambar 21 *Confusion Matrix*

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

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 22 *f1, Accuracy, Precision, & Recall Score*

Kemudian dalam evaluasi terhadap algoritma *Support Vector Machine* dengan *confusion matrix* didapatkan *F1 score, accuracy,*

precision, dan *recall*. Pada Gambar 22 f1 Score yang 18 peroleh 0.78, dimana hasil ini didapatkan dari perbandingan rata-rata *precision* dan *recall* yang dibobotkan dengan rumus : $2 \times (\text{recall} \times \text{precision}) / (\text{recall} + \text{precision})$, sehingga dihasilkan 0.78 atau 78%. *Accuracy* 15 score mendapatkan nilai sebesar 0.76, *accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data, yang dihitung dengan rumus : $(\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$, sehingga menghasilkan 0.76 atau 76%. Kemudian pada *precision score* 12 didapatkan nilai sebesar 0.75, yang merupakan rasio prediksi benar positif dibandingkan dengan hasil prediksi positif, yang dihitung dengan rumus : $(\text{TP}) / (\text{TP} + \text{FP})$, sehingga menghasilkan 0.75 atau 75% nilai dari *precision*. Serta, pada 15 *recall score* diperoleh nilai sebesar 0.82, yang merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif, yang dihitung dengan rumus : $(\text{TP}) / (\text{TP} + \text{FN})$, sehingga menghasilkan nilai 0.82 atau 82% *recall score*.

Visualisasi

Pada tahap ini data yang telah diproses hingga validasi akan divisualisasikan dengan penyajian yaitu dalam dua bentuk penyajian *Wordcloud* dan *Pie Chart*. Visualisasi data dengan *wordcloud* menghasilkan kata yang sering muncul dalam data *tweet* yang telah dianalisis. Kemudian, visualisasi dengan *pie chart* menyajikan perbandingan persentase sentimen positif dan negatif.

Wordcloud

Penggunaan *wordcloud* akan memvisualisasikan kumpulan kata yang sering muncul dalam data set yang telah dianalisis sebelumnya, perbedaan *font zise* pada *wordcloud* menunjukkan frekuensi kemunculan kata pada data set. Hasil visualisasi *wordcloud* dapat dilihat pada gambar 23 untuk sentimen positif dan gambar 24 untuk sentimen negatif.



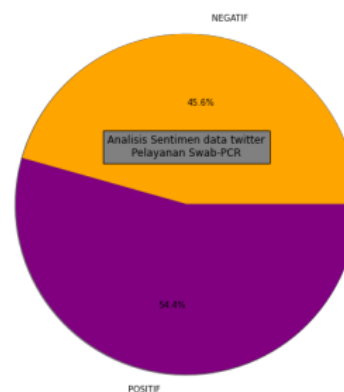
Gambar 23 Sentimen Negatif



Gambar 24 Sentimen Positif

Pie Chart

Penggunaan *pie chart* atau diagram lingkaran akan memvisualisasikan data set dalam bentuk presentase hasil analisis. Persentase sentimen pada data *tweet* pelayanan *swab-PCR* sebesar 54,4% untuk positif dan 45,6 % untuk negatif. Hasil penyajian visualisasi menggunakan *pie chart* dapat dilihat ada gambar berikut ini.



Gambar 25 Pie Chart

17

KESIMPULAN DAN SARAN

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

Kesimpulan

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 0.625, *fold* 7 sebesar 0.625, *fold* 8 sebesar 0.75, *fold* 9 sebesar 1, dan *fold* 10 sebesar 0.75, serata didapatkan rata-rata sebesar 0.66.
2. Didapatkan nilai pengujian pada nilai *accuracy* sebesar 76%, *precision* sebesar 75%, dan *recall* sebesar 81%.
3. Berdasarkan visualisasi *wordcloud* didapatkan kata yang paling positif ialah “periksa” dan “sehat”, serta kata yang paling negatif ialah “hasil” dan “gratis”.
4. Persentase sentimen positif sebesar 54.4% dan persentase sentimen negatif sebesar 45.6% yang divisualisasikan melalui *pie chart* berdasarkan data set dengan *keyword* “pelayanan *swab pcr*”.

Saran

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

- Anoroso, 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/>
- Friyadie. (2016). Penerapan Metode Simple Additive Weight (SAW) dalam Sistem Pendukung Keputusan Promosi Kenaikan Jabatan. *Jurnal Pilar Nusa Mandiri, XII*(1), 37=45.
- Hakim, D. R., Wahyudin, A., & Thomas, P. (2016). Peran Soft Skill Dalam Memediasi Pengaruh Prestasi Belajar dan Aktivitas Berorganisasi Terhadap Daya Saing Mahasiswa Pendidikan Ekonomi Universitas Kuningan. *Journal of Economic Education*, 5(2), 154-167.
- Helilintar, R., Winarno, W. W., & Fatta, H. A. (2016). Penerapan Metode SAW dan Fuzzy Dalam Sistem Pendukung Keputusan Penerimaan Beasiswa. *Citex Journal*, 3(2), 89-101.
- Hidayat, F. K., & Wahyuni, S. N. (2019). Pendeteksian Minat dan Bakat Menggunakan Metode RIASEC. *Indonesian Journal of Business Intelligence*, II(1), 32-39.
- 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/>
- Loelianto, I., Sofyan, M., & Angriani, H. (2020). PLEMETASI TEORI NB PADA CALON MAHASISWA BARU STMIK KHARISMA MAKASSAR. *SINTECH JOURNAL*, 113.
- Makarim, d. F. (2020, October 20). Perbedaan *pcr* dan *antigen* Retrieved from halodoc: <https://www.halodoc.com/>
- Mufizar, T. (2015). Decision Support System Pemilihan Dosen Berprestasi dengan Metode Simple Additive Weighting. *CSRID Journal*, 7(3), 155-166
- Mustikawati, R. I., Nugroho, M. A., Setyorini, D., Yushita, A. N., & Timur, R. P. (2016). Analisis Kebutuhan Soft Skill dalam Mendukung Karir Alumni Akuntansi. *Jurnal*

- 6 Pendidikan Akuntansi Indonesia, XIV(2), 13-20.
- Pratiwi, I. P., Ferdinandus, F., & Limantara, A. D. (2019). Sistem Pendukung Keputusan Penerima Program Keluarga Harapan (PKH) Menggunakan Metode Simple Additive Weighting. *Jurnal Teknik Informatika, Sistem Informasi, dan Ilmu Komputer*, 8(2), 182-195.
- 16 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.
- 6 Supriyatna, A. (2018). Metode Extreme Programming Pada Pembangunan Web Aplikasi Seleksi Peserta Pelatihan Kerja. *Jurnal Teknik Informatika*, 11(1), 26-38.
- Suranto, & Rusdianti, F. (2018). Pengalaman Berorganisasi dalam Membentuk Soft Skill Mahasiswa. *Jurnal Pendidikan dan Ilmu Sosial*, 28(1), 58-65.
- 7 Surya, C. (2015). Sistem Pendukung Keputusan Rekomendasi Penerima Beasiswa Menggunakan Fuzzy Multi Attribut Decision Making (FMADM) dan Simple Additive Weighting (SAW). *Jurnal Rekayasa*
- 20 *Elektrika*, 11(4), 149-156.
- Taufik, & Pamungkas, S. (2018). ANALISIS SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) .

SVM

ORIGINALITY REPORT

27%

SIMILARITY INDEX

26%

INTERNET SOURCES

12%

PUBLICATIONS

14%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur Student Paper	5%
2	primayahospital.com Internet Source	2%
3	jurnal.dinamika.ac.id Internet Source	2%
4	dspace.uii.ac.id Internet Source	1%
5	123dok.com Internet Source	1%
6	ejournal.nusamandiri.ac.id Internet Source	1%
7	jurnal.stmik-amik-riau.ac.id Internet Source	1%
8	j-ptiik.ub.ac.id Internet Source	1%
9	journal.unusida.ac.id Internet Source	1%
10	ojs.ekuitas.ac.id Internet Source	1%

11	publishing-widyagama.ac.id Internet Source	1 %
12	Submitted to Universitas Brawijaya Student Paper	1 %
13	eprints.mdp.ac.id Internet Source	1 %
14	www.syekhnurjati.ac.id Internet Source	1 %
15	Submitted to Sriwijaya University Student Paper	1 %
16	ejournal.bsi.ac.id Internet Source	1 %
17	text-id.123dok.com Internet Source	1 %
18	medium.com Internet Source	1 %
19	journal.walisongo.ac.id Internet Source	<1 %
20	core.ac.uk Internet Source	<1 %
21	Submitted to Universitas Pelita Harapan Student Paper	<1 %
22	journal.uinjkt.ac.id Internet Source	<1 %
23	jurnal.untan.ac.id Internet Source	

<1 %

24 pt.scribd.com
Internet Source

<1 %

25 Oke Dwiraswati, Kemal Nazaruddin Siregar.
"ANALISIS SENTIMEN PADA TWITTER
TERHADAP PENGGUNAAN ANTIBIOTIK DI
INDONESIA DENGAN NAIVE BAYES
CLASSIFIER", Media Informasi, 2019
Publication

<1 %

26 jnk.phb.ac.id
Internet Source

<1 %

27 Submitted to Universitas Jember
Student Paper

<1 %

28 repository.dinamika.ac.id
Internet Source

<1 %

29 adoc.pub
Internet Source

<1 %

30 cogito.unklab.ac.id
Internet Source

<1 %

31 repository.uinsu.ac.id
Internet Source

<1 %

32 staf.cs.ui.ac.id
Internet Source

<1 %

33 swara.tunaiku.com
Internet Source

<1 %

34 www.e-sciencecentral.org <1 %
Internet Source

35 ejournal.stiki-indonesia.ac.id <1 %
Internet Source

36 www.slideshare.net <1 %
Internet Source

37 Hendry Cipta Husada, Adi Suryaputra
Paramita. "Analisis Sentimen Pada
Maskapai Penerbangan di Platform Twitter
Menggunakan Algoritma Support Vector
Machine (SVM)", Teknika, 2021
Publication

38 Luluk Suryani, Kasmi Edy.
"PENGEMBANGAN APLIKASI "LOST &
FOUND" BERBASIS ANDROID DENGAN
MENGUNAKAN METODE TERM
FREQUENCY – INVERSE DOCUMENT
FREQUENCY (TF-IDF) DAN COSINE
SIMILARITY", Electro Luceat, 2020
Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography On