

Sentimen Analisis Penerapan *Social Distancing* Menggunakan *Feature Selection* Pada Algoritma *Support Vector Machine*

Siti Nurhasanah Nugraha¹, Tri Rivanie², Sri Rahayu³, Windu Gata⁴, Rangga Pebrianto⁵

^{1,2,3,4,5}STMIK Nusa Mandiri

¹e-mail: 14002394@nusamandiri.ac.id

²e-mail: 14002326@nusamandiri.ac.id

³e-mail: srirahayu.rry@nusamandiri.ac.id

⁴e-mail: windu@nusamandiri.ac.id

⁵e-mail: 14002396@nusamandiri.ac.id

Diterima	Direvisi	Disetujui
09-05-2020	14-07-2020	29-07-2020

Abstrak - "*Social Distancing*" atau dalam kata lain mengurangi interaksi sosial dengan orang lain sangat berpotensi dalam menyelamatkan kehidupan jutaan orang di masa *pandemic COVID-19*. Anjuran pemerintah kepada masyarakat untuk penerapan *social distancing* tak lepas dari pendapat atau opini pengguna sosial media yang salah satunya disuarakan melalui *twitter*. Salah satu metode atau teknik untuk mengelompokkan kategori opini atau pendapat dari pengguna sosial media adalah *sentiment analyst*. Penelitian ini menggunakan *dataset* yang di *crawling* dari *twitter* dengan kata kunci "*Social Distancing*". Data hasil *crawling* tersebut diolah menggunakan model algoritma SVM dengan penambahan *feature selection Weight by Correlation*. Penelitian ini membandingkan hasil *cross validation* algoritma SVM tanpa *feature selection* dan algoritma SVM dengan *feature selection*. Hasil *cross validation* SVM tanpa *feature selection* menunjukkan nilai *accuracy* sebesar 67,00% dan nilai AUC sebesar 0,709. Sedangkan hasil *cross validation* algoritma SVM dengan *feature selection* menunjukkan nilai *accuracy* sebesar 70,33% dan nilai AUC sebesar 0,838. Dari kedua model tersebut diketahui bahwa penggunaan *feature selection Weight by Correlation* dapat meningkatkan nilai *Accuracy* dan nilai AUC. Sehingga dapat disimpulkan bahwa penambahan *feature selection Weight by Correlation* sangat baik apabila digunakan pada algoritma *Support Vector Machine (SVM)* karena dapat menghasilkan nilai akurasi dan nilai AUC yang lebih tinggi.

Kata Kunci: *sentimen analisis, social distancing, feature selection, SVM*.

Abstract - "*Social Distancing*" or in other words reducing social interaction with others has the potential to save the lives of millions of people during the *COVID-19 pandemic*. The government's suggestion to the public for the implementation of *social distancing* cannot be separated from the opinions or opinions of social media users, one of which is voiced through *Twitter*. One method or technique for classifying opinion categories or opinions from social media users is *sentiment analyst*. This study uses a *dataset* that is *crawled* from *Twitter* with the keyword "*Social Distancing*". The *crawling* data is processed using the *SVM* algorithm model with the addition of the *Weight by Correlation selection feature*. This research compares the results of *cross validation* of *SVM* algorithm without *feature selection* and *SVM* algorithm with *feature selection*. The results of *cross validation* *SVM* without *feature selection* show an *accuracy* value of 67.00% and *AUC* value of 0.709. While the results of *cross validation* of *SVM* algorithm with *feature selection* show an *accuracy* value of 70.33% and *AUC* value of 0.838. From the two models it is known that the use of the *Weight by Correlation selection feature* can increase the value of *Accuracy* and *AUC* values. So it can be concluded that the addition of *Weight by Correlation feature selection* is very good when used in the *Support Vector Machine (SVM)* algorithm because it can produce higher *accuracy* values and *AUC* values.

Keywords: *sentiment analysis, social distancing, feature selection, SVM*.

PENDAHULUAN

"*Social Distancing*" atau dalam kata lain mengurangi interaksi sosial dengan orang lain sangat berpotensi dalam menyelamatkan kehidupan jutaan orang di masa *pandemic COVID-19* ini (Ferguson et

al., 2020). Pemerintah di seluruh dunia telah menerapkan berbagai tingkat *social distancing*, akan tetapi kepatuhan dari setiap masing-masing individu itulah yang paling penting (Anderson, Heesterbeek, Klinkenberg, & Hollingsworth, 2020). Banyak negara yang mengadopsi langkah-langkah yang

disebut *social distancing* atau *physical distancing*, menutup sekolah dan tempat kerja serta membatasi pertemuan, salah satunya di Negara Indonesia. Tujuan diberlakukannya strategi ini adalah untuk memperlambat penyebaran infeksi dan mengurangi intensitas epidemi (Anderson et al., 2020).

Perkembangan media yang begitu pesat memunculkan banyak media *online* dari media berita sampai media sosial. Media sosial sudah begitu banyak, mulai dari *Facebook*, *Twitter*, *Path*, *Instagram*, *Google+*, *Tumblr*, *Linkedin* dan masih banyak lagi. Media sosial khususnya *twitter* sekarang ini menjadi perangkat komunikasi yang sangat populer di kalangan pengguna internet (Buntoro, 2017).

Di masa pandemic *COVID-19* ini banyak warga negara Indonesia yang menyuarakan pendapat atau opini mengenai penerapan *social distancing* di Indonesia pada sosial media, salah satunya adalah dengan melalui *twitter*. Penerapan *social distancing* di Indonesia tidak lepas dari opini positif bahkan opini negatif yang dilontarkan oleh sebagian pengguna *twitter*. Mengolah opini dari pengguna sosial media bukanlah hal yang mudah, dengan jumlah opini yang terdapat pada sosial media yang begitu banyak untuk diproses secara manual. Maka dari itu untuk mengelompokkan apakah opini tersebut termasuk kategori positif atau negatif, diperlukan sebuah metode atau teknik khusus. Melalui *twit* yang di *share* oleh pengguna ini lah yang bisa diambil dan dijadikan analisis sentimen mengenai penerapan *social distancing* di Indonesia.

Machine Learning saat ini sedang sangat ramai dibicarakan banyak orang, dengan kemampuannya untuk bisa menjalankan kemampuan manusia oleh sebuah mesin. *Machine learning* adalah bidang studi yang memberikan kemampuan pada program komputer untuk belajar tanpa secara eksplisit diprogram (Fikriya, Irawan, & Soetrisno., n.d.). Salah satu kemampuan yang dikembangkan oleh praktisi *machine learning* adalah *sentiment analyst*. *Sentiment Analyst* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik (Saputra, Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain, 2019). Analisis sentimen juga diartikan sebagai proses memahami, mengekstrak dan mengolah data tekstual secara otomatis dengan tujuan untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Dengan dilakukannya penelitian ini diharapkan dapat diketahui analisis sentimen para pengguna *twitter* terhadap penerapan *social distancing* di Indonesia.

Penelitian ini menggunakan *dataset* yang di *crawling* dari *twitter* dengan kata kunci "*Social Distancing*". Pada penelitian ini menggunakan

algoritma *Support Vector Machine* (SVM) dengan penambahan *feature selection* yakni *weight by correlation* yang merupakan salah satu metode yang digunakan untuk menaikkan nilai akurasi dari model *machine learning* yang dikembangkan. *Weight by Correlation* sendiri artinya pembobotan pada atribut dengan cara menghubungkan (korelasi) satu attribute dengan atribut lainnya (Documentation, 2019). *Weight by Correlation* juga merupakan *feature selection* dalam analisis sentimen yang diletakkan sebelum digunakannya algoritma sebagai model. Adapun optimasi yang diletakkan pada akhir model analisis sentimen, salah satunya adalah *Particle Swarm Optimization* (PSO) namun PSO memiliki ketergantungan yang sensitif pada parameter yang digunakan (Achyani, 2018).

Pada penelitian ini akan membandingkan apakah metode SVM saja dengan SVM yang menggunakan tambahan *feature selection* dapat meningkatkan nilai *accuracy* dan membandingkan nilai AUC dari kedua metode tersebut. SVM telah menjadi metode klasifikasi dan regresi yang sering digunakan untuk masalah linear dan nonlinear. Kelebihan dari algoritma ini adalah dari kemampuannya untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang diperlukan (Santoso, Gata, & Paryanti, 2019). Efektivitas SVM sangat dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik (Haddi, Liu, & Shi, 2013). Banyak peneliti yang telah membuktikan bahwa SVM merupakan metode yang paling akurat untuk teks klasifikasi.

Penelitian sebelumnya yang membahas mengenai optimasi pada algoritma klasifikasi dengan menggunakan metode yang sama dengan hasil *accuracy* sebesar 81,18% (Santoso et al., 2019). Penelitian mengenai analisis sentimen juga telah dilakukan untuk *review* kosmetik menggunakan algoritma SVM dengan penambahan *feature selection partical swarm optimazion* (PSO) yang menghasilkan peningkatan nilai *accuracy* dari 89,00% menjadi 97,00% (Kristiyanti, 2015). Penelitian yang membahas penggunaan algoritma SVM berbasis algoritma genetika untuk analisis sentimen pada data *twitter* menghasilkan peningkatan nilai akurasi dari 83,50% menjadi 91,00% (Putri, 2015).

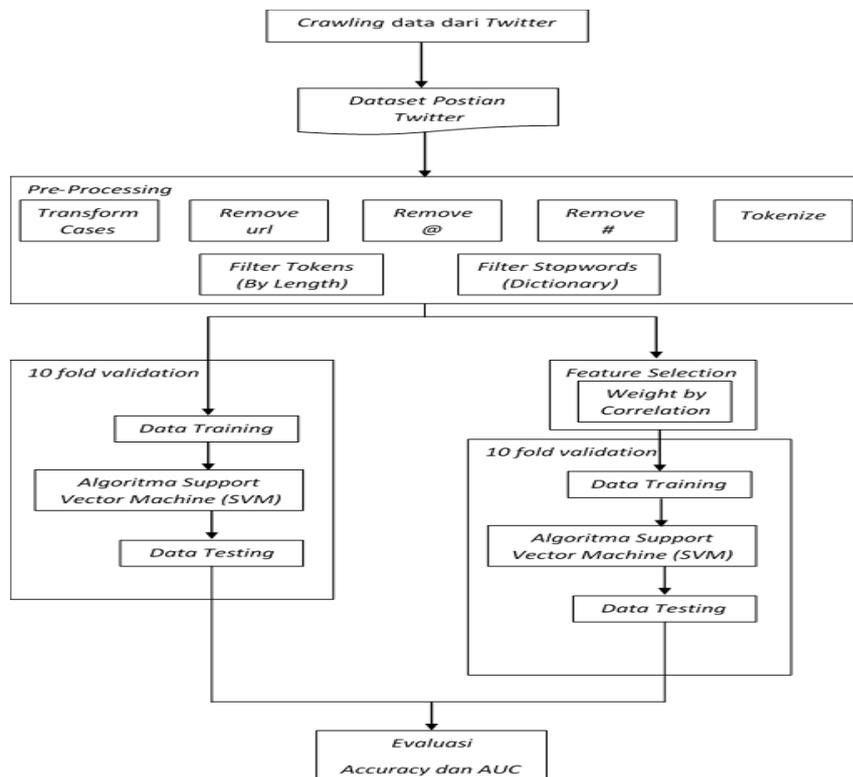
Penggunaan *feature selection* tidak hanya digunakan pada algoritma SVM, ada juga penelitian yang menggunakan *feature selection* pada algoritma *Naive Bayes* untuk analisis sentimen *movie review* yang menghasilkan *accuracy* sebesar 95,70% (Andilala, 2016). Dapat dibuktikan bahwa dengan penggunaan *feature selection* dapat meningkatkan nilai *accuracy* (Kurniawati & Pardede, 2018). Dengan adanya penelitian ini diharapkan akan menghasilkan peningkatan yang signifikan seperti

pada penelitian-penelitian sebelumnya yang sudah dilakukan.

METODOLOGI PENELITIAN

Dalam konteks penelitian, metode yang dilakukan mengacu pada pemecahan masalah yang meliputi

pengumpulan data, perumusan hipotesis, pengujian hipotesis, menafsirkan hasil pembahasan dan penarikan kesimpulan (Achyani, 2018). Metode penelitian yang digunakan pada penelitian ini dapat dilihat pada gambar dibawah ini.



Sumber: Penelitian (2020)

Gambar 1. Model Penelitian

1. Pengambilan Data

Proses pengambilan data pada penelitian ini dilakukan dengan *crawling* data dari *twitter*, kemudian dikelola dengan bantuan aplikasi *rapidminer* untuk selanjutnya masuk tahap *preprocessing*.

2. Pre-Processing Data

Sebelum *dataset* dimasukkan kedalam model maka dilakukan *pre-processing data* terlebih dahulu agar dapat diolah pada saat membuat pemodelan. Tahap ini meliputi kegiatan membangun data dan juga membersihkan data agar siap untuk dikelola pada tahap selanjutnya. Proses pertama *transform cases* yaitu untuk mengubah bentuk huruf menjadi kapital yang ada pada *text* menjadi huruf kecil semua (*lowercase*). Hal ini dilakukan untuk penyeragaman bentuk huruf, serta untuk meminimalisir kesalahan pada saat proses *tokenize*. Kemudian dilanjutkan pada proses *remove URL* atau menghilangkan

Uniform Resouce Locator (URL) yang ada pada *dataset* yang dihasilkan dari proses *crawling* data. Kemudian dilanjutkan pada proses *annotation removal*, yaitu menghapus tanda *annonate* (@) dan teks yang mengikutinya. Selanjutnya proses *remove hashtag* (#), yaitu menghapus tanda (#) yang ada pada teks.

Kemudian selanjutnya adalah *tokenize* yaitu proses yang akan membelah kata-kata yang ada dalam *dataset* menjadi terpisah, bukan merupakan kalimat lagi. Kemudian kata-kata tersebut akan disebut dengan istilah fitur. Kemudian dilanjutkan pada proses *Filter Tokens (by Length)*, yang akan menghilangkan kata-kata dengan panjang karakter tertentu. Karena biasanya kata yang hanya memiliki 2 karakter tidak mempunyai arti. Setelah itu dilakukan proses *filter stopwords* yaitu membuang kata-kata yang tidak mempunyai arti tepat pada sentimen analisis, biasanya berupa kata

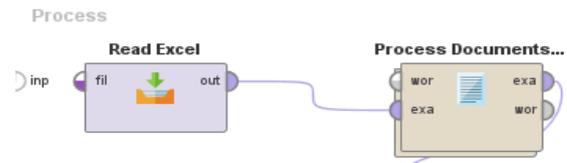
keterangan dan kata sambung. Proses terakhir adalah *Feature Seleccion (Weight by Correlation)* yaitu proses menghitung bobot untuk masing-masing fitur (*text*) yang ada dalam *dataset*. Kemudian dilakukan pemilihan fitur dengan kriteria bobot terurut dari terkecil ke terbesar.

3. *Modelling*

Pada tahap ini dilakukan *10 fold cross validation* untuk membagi *dataset* menjadi 10 bagian dimana 1 diantara bagian lainnya menjadi data *testing*, sedangkan data lainnya menjadi data *training*. Kemudian dimasukkan kedalam model algoritma SVM.

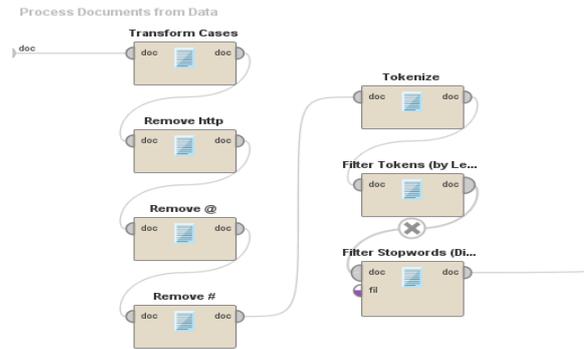
4. *Evaluation*

Setelah tahapan modelling selesai maka tahap selanjutnya adalah melakukan evaluasi terhadap hasil dari pemodelan tersebut. Evaluasi yang dilakukan yaitu dengan membandingkan dua hasil dari pemodelan yang berupa *accuracy*, *precision*, *recall* maupun AUC antara model dengan *Feature Seleccion (Weight by Correlation)* maupun tanpa *feature selection*.



Sumber: Hasil Penelitian (2020)

Gambar 3. Pengambilan Data setelah *Cleansing*



Sumber: Hasil Penelitian (2020)

Gambar 4. Operator Process Document

Operator yang terdapat dalam *process document* adalah operator *transform cases*, *tokenize (remove url, remove #, remove @)*, *tokenize*, *filter tokens* dan *filter stopwords*. Adapun langkah-langkahnya sebagai berikut:

a. *Transform Case*

Pada penelitian ini menggunakan fitur *lowercase* untuk mengubah bentuk huruf kapital yang ada pada *text* menjadi huruf kecil semua. Perbedaannya seperti pada tabel berikut.

Tabel 1. Hasil *Transform Cases*

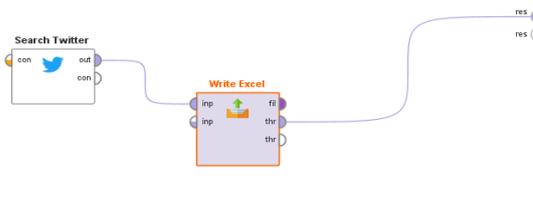
Text Asli	<i>Transform Cases</i>
Terus waspada, jaga disiplin dalam social distancing, Jangan mudik !!!!!!!!!!!!!,Kasus Infeksi Corona Tanpa Gejala di China Melonjak https://t.co/zmlTgsc1A4	terus waspada, jaga disiplin dalam social distancing, jangan mudik !!!!!!!!!!!!!,kasus infeksi corona tanpa gejala di china melonjak https://t.co/zmltgsc1a4
FiqhTabayyun: Mcm ni ke babi punya PKP? Dok menjalar kesana dan kesini. Macam esok lusa nak kiamat dah takut tak jumpa org2 ni semua....	fiqhtabayyun: mcm ni ke babi punya pkp? dok menjalar kesana dan kesini. macam esok lusa nak kiamat dah takut tak jumpa org2 ni semua....
@xxnryo SAMA WOI SELAMA SOCIAL DISTANCING DUIT JUGA IKUTAN DISTANCING??	@xxnryo sama woi selama social distancing duit juga ikutan distancing??

Sumber: Hasil Penelitian (2020)

HASIL DAN PEMBAHASAN

1. Pengambilan *Dataset*

Pengambilan *dataset* pada penelitian ini diambil dari proses *crawling* dari sosial media *twitter* pada tools *rapidminer* dengan menggunakan operator *search twitter* dengan *query* “*Social Distancing*”. Kemudian *dataset* tersebut disimpan kedalam format *excel* untuk dilakukan proses *cleansing* yang terdiri dari *remove duplicate data* dan pelabelan secara manual. Proses pengambilan *dataset* seperti pada gambar berikut.



Sumber: Hasil Penelitian (2020)

Gambar 2. Proses *Crawling* Data

2. Proses *Pre-Processing*

Setelah *crawling* data dari *twitter* dan data sudah di *cleansing*, kemudian dilanjutkan dengan tahap *pre-processing*. Data tersebut tidak bisa langsung dimasukkan dalam pengolahan untuk sentimen analisis, maka dilanjutkan dengan tahapan *pre-processing*.

b. *Remove URL*

Pada operator ini akan menghilangkan URL yang ada pada *text*. Berikut perbedaan antara *text* yang belum di *remove url* dan yang sudah di *remove url*.

Tabel 2. Hasil *Remove URL*

Text Asli	<i>Remove URL</i>
Terus waspada, jaga disiplin dalam social distancing, Jangan mudik !!!!!!!...,Kasus Infeksi Corona Tanpa Gejala di China Melonjak https://t.co/zmlTgsc1A4	terus waspada, jaga disiplin dalam social distancing, jangan mudik !!!!!!!...,kasus infeksi corona tanpa gejala di china melonjak
Ambedkar peru meeda social distancing ni ...Photo kosam mask ni pakkaki pettina TRS MLA bhupal reddy garu ????.Oka maata cheppandi @KTRTRS garu https://t.co/LDIwyk0Ny0	ambedkar peru meeda social distancing ni ...photo kosam mask ni pakkaki pettina trs mla bhupal reddy garu ????.oka maata cheppandi @ktrtrs garu
Orangorang pda stres akibat social distancing aoa gmn sih nih ? Pecinta kucing tapi maunya yg ras dasar nga ada otak https://t.co/4Fau4B966y	orangorang pda stres akibat social distancing aoa gmn sih nih ? Pecinta kucing tapi maunya yg ras dasar nga ada otak

Sumber: Hasil Penelitian (2020)

c. *Remove Annotation (@)*

Pada operator ini akan menghilangkan tanda *annotation* (@) yang terdapat pada *text*. Perbedaan dari sebelum proses *annotation removal* dan setelah proses *annotation removal* dapat dilihat pada tabel berikut.

Tabel 3. Hasil *Remove Annotation (@)*

Text Asli	<i>Remove @</i>
@FOOD_FESS mesesnya ikutan social distancing ya jauh amat jaraknya	mesesnya ikutan social distancing ya jauh amat jaraknya
@impolarbear00 @yudi_bojes @EkoIhsan2 @UtoIjo @wawat_kurniawan @mas_hbsan @CNNIndonesia @detikcom di tugu kujang, bukan di istana bogor, ada pema'luman pelanggaran social distancing ?	jadi, kalau di tugu kujang, bukan di istana bogor, ada pema'luman pelanggaran social distancing ?
@Christian2894 @e100ss @_TNIAL_ mohon bantuan penertiban di bandara juanda sesuai arahan Pak @jokowi untuk social distancing, semoga Allah SWT selalu memberikan kesehatan dan perlindungan pada kita semua ?? mungkin bisa tolong bantu kampanye nya di @BanggaSurabaya bisa	mohon bantuan penertiban di bandara juanda sesuai arahan pak untuk social distancing, semoga allah swt selalu memberikan kesehatan dan perlindungan pada kita semua ?? mungkin bisa tolong bantu kampanye nya di bandara juanda??

tolong bantu kampanye nya di bandara juanda??	
---	--

Sumber: Hasil Penelitian (2020)

d. *Remove Hashtag (#)*

Operator ini akan melakukan penghapusan tanda *hashtag* (#) yang ada pada *text*. Perbedaannya dapat dilihat pada tabel berikut.

Tabel 4. Hasil *Remove Hashtag (#)*

Text Asli	<i>Remove #</i>
Vitasea!* *Foto hanya pemanis dan penyegar! Tetap patuhi anjuran pemerintah untuk Social dan Physical Distancing! Semoga Pandemi ini cepat berlalu agar kita bisa kembali menjelajah! #dirumahaja #fotodibaliklayar... https://t.co/2Xuhm9PXLp	vitasea!* *foto hanya pemanis dan penyegar! tetap patuhi anjuran pemerintah untuk social dan physical distancing! semoga pandemi ini cepat berlalu agar kita bisa kembali menjelajah!
Social distancing. Kita rakyat mesti berlatih semandirimumkin. Jika pun nanti kita dihadapkan dengan pemerintahan yang benar benar gatot, maka kita rakyat bisa berdikari. #StayAtHome #cepatsembuhbumiku #cepatmembaikibupertiwiku	social distancing. kita rakyat mesti berlatih semandirimumkin. jika pun nanti kita dihadapkan dengan pemerintahan yang benar benar gatot, maka kita rakyat bisa berdikari.
Katanya social distancing, malah bagi ² sembako di Istana Bogor. Giliran dikritik, yg ngritik ditangkep. Yg mau mesoh ² waktu dan tempat dipersilahkan. #SaveAliBaharsyah #KhilafahAjaranIslam #BebaskanAliBaharsyah #KhilafahAjaranAswaja #StopKriminalisasiIslam https://t.co/td71RSsa5D	katanya social distancing, malah bagi ² sembako di istana bogor. giliran dikritik, yg ngritik ditangkep. yg mau mesoh ² waktu dan tempat dipersilahkan.

Sumber: Hasil Penelitian (2020)

e. *Tokenize*

Proses ini akan menghilangkan tanda baca, angka, simbol, karakter khusus atau apapun yang bukan huruf. Perbedaannya dapat dilihat pada tabel dibawah ini.

Tabel 5. Hasil *Tokenize*

Text Asli	<i>Tokenize</i>
Terus waspada, jaga disiplin dalam social distancing, Jangan mudik !!!!!!!!!,Kasus Infeksi Corona Tanpa Gejala di China Melonjak https://t.co/zmlTgsc1A4	terus waspada jaga disiplin dalam social distancing jangan mudik kasus infeksi corona tanpa gejala di china melonjak
Ambedkar peru meeda social distancing ni ...Photo kosam mask ni pakkaki pettina TRS MLA bhupal reddy garu ????.Oka maata cheppandi @KTRTRS garu https://t.co/LDIwyk0Ny0	ambedkar peru meeda social distancing ni photo kosam mask ni pakkaki pettina trs ml bhupal reddy garu oka maata cheppandi garu
@impolarbear00 @yudi_bojes @EkoIhsan2 @UtoIjo @wawat_kurniawan @mas_hbsan @CNNIndonesia @detikcom @tvOneNews Jadi, kalau di tugu kujang, bukan di istana bogor, ada pema'luman pelanggaran social distancing ?	jadi kalau di tugu kujang bukan di istana bogor ada pemaluman pelanggaran social distancing

Sumber: Hasil Penelitian (2020)

f. *Filter Tokens (By Length)*

Pada penelitian ini panjang minimum karakter yang digunakan adalah 4 karakter dengan panjang maksimum 25 karakter. Artinya kata dengan panjang kurang dari 4 karakter dan lebih dari 25 karakter akan dihilangkan. Berikut ini hasil dari proses *filter tokens (by length)*.

Tabel 6. Hasil *Filter Tokens (by Length)*

Text Asli	<i>Filter Tokens (By Length)</i>
Terus waspada, jaga disiplin dalam social distancing, Jangan mudik !!!!!!!!!,Kasus Infeksi Corona Tanpa Gejala di China Melonjak https://t.co/zmlTgsc1A4	terus waspada jaga disiplin dalam social distancing jangan mudik kasus infeksi corona tanpa gejala china melonjak
Ambedkar peru meeda social distancing ni ...Photo kosam mask ni pakkaki pettina TRS MLA bhupal reddy garu ????.Oka maata cheppandi @KTRTRS garu https://t.co/LDIwyk0Ny0	ambedkar peru meeda social distancing photo kosam mask pakkaki pettina bhupal reddy garu maata cheppandi garu
@impolarbear00 @yudi_bojes @EkoIhsan2 @UtoIjo @wawat_kurniawan @mas_hbsan @CNNIndonesia @detikcom @tvOneNews Jadi, kalau di tugu kujang, bukan di istana bogor, ada pema'luman pelanggaran social distancing ?	jadi kalau tugu kujang bukan istana bogor pemaluman pelanggaran social distancing

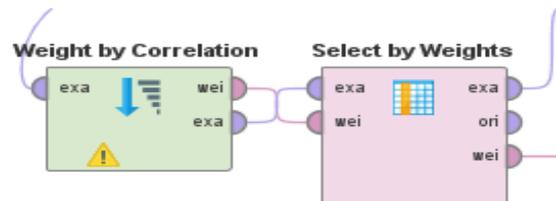
Sumber: Hasil Penelitian (2020)

g. *Stopword Removal*

Selanjutnya adalah penggunaan operator *Stopword Removal (by Directory)* yang berfungsi untuk menghilangkan kata-kata yang tidak memiliki makna.

3. *Feature Selection (Weight by Correlation)*

Langkah selanjutnya yaitu penggunaan operator *Weight by Correlation* yang berfungsi untuk menghitung bobot berdasarkan korelasi antara satu fitur dengan fitur lain yang terdapat pada *dataset* dengan mempertimbangkan nilai label. Setelah dilakukan penghitungan bobot maka data diurutkan dari nilai terkecil ke nilai terbesar (*sort ascending*). Kemudian dilakukan pengambilan fitur yang mempunyai urutan 1 sampai dengan 600. Hal ini dilakukan dengan menggunakan menggunakan operator *select by weight* dengan parameter *weight relation = top k* dengan nilai $k = 600$. Penggunaan operator tersebut dapat dilihat pada gambar dibawah ini.



Sumber: Hasil Penelitian (2020)

Gambar 5. *Feature Selection (Weight by Correlation)*

Contoh atribut/fitur yang telah di bobotkan dan diurutkan mulai dari bobot terkecil ke terbesar (*ascending*) dapat dilihat pada gambar berikut.

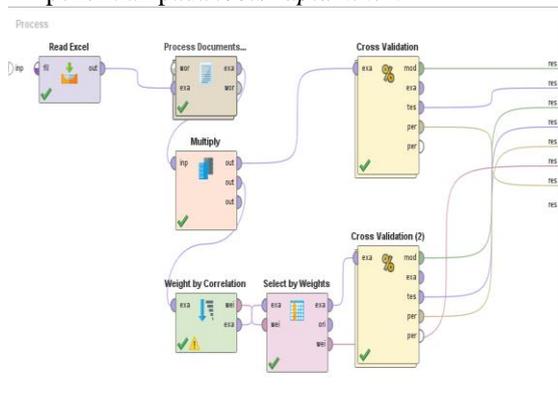
attribute	weight
mask	0.000
distancin	0.001
knyer	0.001
sosial	0.001
allyhapi	0.001
jawaban...	0.001
hoax	0.001
sepi	0.002
kesana	0.002
dirumah...	0.002
membut...	0.002
melulu	0.002
menerap...	0.003

Sumber: Hasil Penelitian (2020)

Gambar 6. Atribut Hasil Pembobotan

4. Modelling

Model dalam penelitian ini digambarkan seperti pada gambar dibawah ini yang merupakan keseluruhan operator yang digunakan dalam penelitian pada *tools rapidminer*.



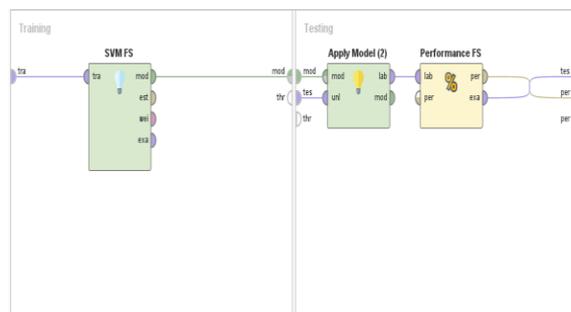
Sumber: Hasil Penelitian (2020)

Gambar 7. Model Penelitian pada *Rapidminer*

Pada pemodelan diatas dapat diketahui bahwa penggunaan algoritma SVM yang ada dalam operator *Cross Validation* terdapat 2 buah. Perbedaannya adalah salah satu tanpa dilakukan pembobotan dengan korelasi (hasil dari *pre-processing data* langsung ke *Cross Validation*) dan yang satunya menggunakan pembobotan korelasi. Untuk membedakan kedua proses tersebut digunakan operator *Multiply* untuk membagi *exampleset output* dari *processing data* masuk ke *cross validation* untuk diproses dengan algoritma SVM dan dilakukan pengukuran evaluasi.

5. Evaluasi

Untuk melakukan evaluasi dalam penelitian ini menggunakan *Accuracy* dan *AUC (Area Under Curve)*. Hasil dari SVM tanpa dan dengan *feature selection* algoritma akan dibandingkan untuk mengukur seberapa besar peningkatan yang dihasilkan. Untuk mendapatkan nilai yang akan dievaluasi maka digunakan *10 fold cross validation* terhadap data yang telah dihitung pembobotannya, dapat dilihat pada gambar dibawah.



Sumber: Hasil Penelitian (2020)

Gambar 8. *Cross Validation*

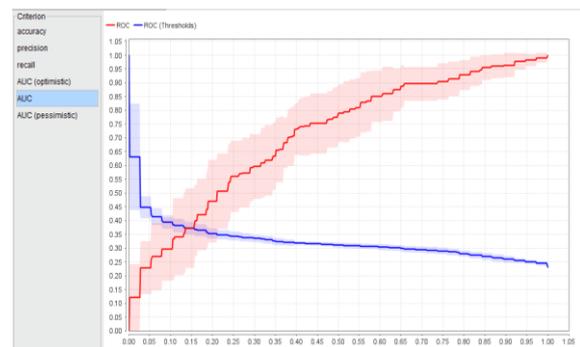
Dari tahapan yang telah dilakukan maka dihasilkan nilai *Accuracy* dan *AUC* dari algoritma SVM tanpa *feature selection* sebagai berikut.

Tabel 7. Hasil *Accuracy* dan *AUC* tanpa *Feature Selection*

SVM tanpa <i>Feature Selection</i>	
<i>Accuracy</i>	67,00%
<i>AUC</i>	0,709

Sumber: Hasil Penelitian (2020)

Dapat diketahui bahwa *cross validation* SVM tanpa *feature selection* mendapatkan *accuracy* sebesar 67,00% dengan nilai *AUC* sebesar 0,709. Bentuk kurva dari ROC yang dihasilkan dapat dilihat pada gambar dibawah.



Sumber: Hasil Penelitian (2020)

Gambar 9. Kurva ROC tanpa *Feature Selection*

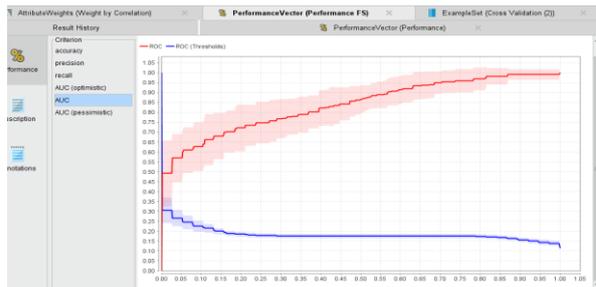
Sedangkan untuk nilai *accuracy* dan *AUC* dari *cross validation* algoritma SVM dengan penambahan *feature selection* dapat dilihat pada tabel berikut.

Tabel 8. Hasil *Accuracy* dan *AUC* dengan *Feature Selection*

SVM dengan <i>Feature Selection</i>	
<i>Accuracy</i>	70,33%
<i>AUC</i>	0,838

Sumber: Hasil Penelitian (2020)

Dari tabel diatas diketahui bahwa hasil *cross validation* algoritma SVM dengan penambahan *feature selection* mendapatkan nilai *Accuracy* 70,33% dan *AUC* 0,838, dengan kurva ROC seperti pada gambar dibawah.



Sumber: Hasil Penelitian (2020)

Gambar 10. Kurva ROC dengan *Feature Selection*

Dari kedua hasil *cross validation* diatas maka dapat diketahui selisih peningkatan dengan penambahan *feature selection* untuk nilai *accuracy* dan AUC. Peningkatan nilai *accuracy* sebesar 3,33% dan nilai AUC sebesar 0,129, dapat dilihat pada tabel dibawah.

Tabel 9. Selisih Nilai *Accuracy* dan AUC

	SVM tanpa <i>feature selection</i>	SVM dengan <i>feature selection</i>	Selisih
<i>Accuracy</i>	67,00%	70,33%	3,33%
AUC	0,709	0,838	0,129

Sumber: Hasil Penelitian (2020)

KESIMPULAN

Dari penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa analisis sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya pengguna *twitter* terhadap penerapan *social distancing* dimasa pandemic ini. Dengan tujuan untuk membantu masyarakat menentukan sentimen yang terdapat pada *twit* opini yang ada di *twitter*. Penelitian analisis sentimen ini dilakukan dengan menggunakan model algoritma SVM tanpa *feature selection* dan membandingkannya dengan model SVM dengan *feature selection*. Dari kedua model tersebut diketahui bahwa penggunaan *feature selection Weight by Correlation* dapat meningkatkan nilai *Accuracy* dan nilai AUC. Sehingga dapat disimpulkan bahwa penambahan *feature selection Weight by Corellation* sangat baik apabila digunakan pada algoritma *Support Vector Machine (SVM)* karena dapat menghasilkan nilai akurasi dan nilai AUC yang lebih tinggi. Penelitian selanjutnya disarankan untuk dapat menggunakan *feature selection Weight* lainnya selain *feature selection Weight by Correlation*. Pada penelitian selanjutnya juga dapat dilakukan dengan penambahan jumlah data yang lebih banyak untuk mendapatkan nilai akurasi yang lebih baik.

REFERENSI

Achyani, Y. E. (2018). Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi

Pemasaran Langsung. *Jurnal Informatika*, 5(1), 1–11. <https://doi.org/10.31311/ji.v5i1.2736>

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228), 931–934. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5)

Andilala, A. (2016). Movie Review Sentimen Analisis Dengan Metode Naïve Bayes Base on Feature Selection. *Pseudocode*, 3(1), 1–9. <https://doi.org/10.33369/pseudocode.3.1.1-9>

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal Maret*, 1(1), 32–41. Retrieved from https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf

Documentation, R. (2019). Weight by Correllation. Retrieved April 25, 2020, from https://docs.rapidminer.com/latest/studio/operators/modeling/f%0Aeature_weights/weight_by_correlation.html

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., ... Ghani, A. C. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial.Ac.Uk*, (March), 3–20. <https://doi.org/10.25561/77482>

Fikriya, Z. A., Irawan, M. I., & Soetrisno., S. (n.d.). *Implementasi Extreme Learning Machine untuk Pengenalan Objek Citra Digital*. Retrieved from http://ejurnal.its.ac.id/index.php/sains_seni/article/view/21754

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26–32. <https://doi.org/10.1016/j.procs.2013.05.005>

Kristiyanti, D. A. (2015). ANALISIS SENTIMEN REVIEW PRODUK KOSMETIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN PARTICLE SWARM OPTIMIZATION SEBAGAI.

Kurniawati, I., & Pardede, H. F. (2018). Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis. *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*, 1–5. <https://doi.org/10.1109/ICITSI.2018.8695953>

Putri, D. A. (2015). ALGORITMA SUPPORT

*VECTOR MACHINE BERBASIS
ALGORITMA GENETIKA UNTUK ANALISIS
SENTIMEN PADA.*

- Santoso, I., Gata, W., & Paryanti, A. B. (2019). *Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum. 1(10), 5–11.*
- Saputra, S. A., Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain. (2019). *Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 3(3), 377–382. <https://doi.org/10.29207/resti.v3i3.1118>*