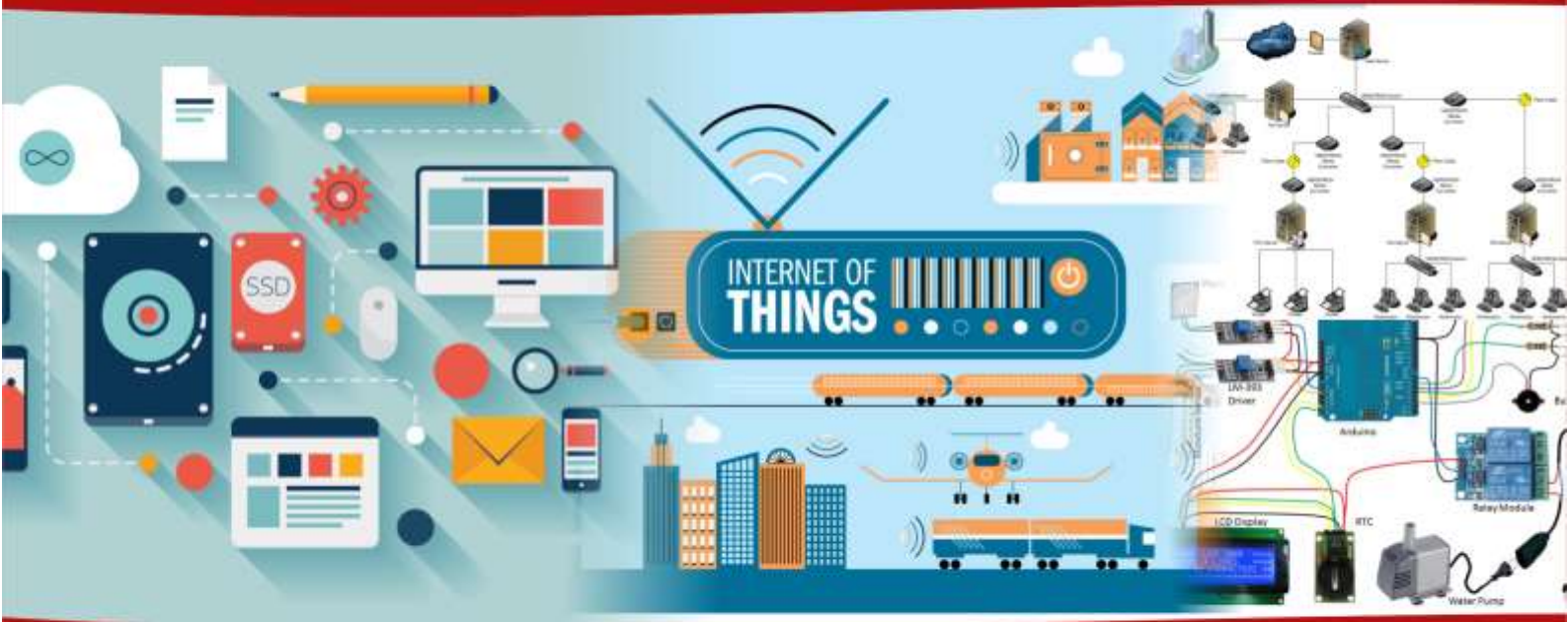
ISSN : 2620-6897 (*Cetak*)
ISSN : 2620-6900 (*Online*)

Volume 3, Nomor 2, November 2020

JIRE

JURNAL INFORMATIKA & REKAYASA ELEKTRONIKA



Diterbitkan Oleh LPPM STMIK Lombok
Jln. Basuki Rahmat No.105 Praya, Lombok Tengah - NTB
e-journal.stmiklombok.ac.id/jire - Telp dan Fax (0370) 654310
email. lppm@stmiklombok.ac.id

DEWAN REDAKSI

Jurnal Manager
Wire Bagye, S.Kom.,M.Kom *(STMIK Lombok, SINTA ID : 5992010)*

Reviewer :
Resad Setyadi.,S.T.,S.Si.,MMSI.,Ph,D (cand) - *Institut Teknologi Telkom Purwokerto*
SCOPUS ID : 57204172534 *SINTA ID* : 6113570
Yesaya Tommy Paulus, S.Kom., MT., Ph.D. - *STMIK Dipanegara Makassar*
SCOPUS ID : 57202829909 *SINTA ID* : 6002004
Dr. Cucut Susanto, S. Kom. MSi. - *STMIK Dipanegara Makassar*
SINTA ID : 6138863
Muhamad Malik Mutoffar, ST., MM., CNSS- *Sekolah Tinggi Teknologi Bandung*
SINTA ID : 6013819
David, M.Cs.,M.Kom - *STMIK Pontianak*
SCOPUS ID : 57200208543 *SINTA ID* : 5977352
Indo Intan, S.T., M.T. *STMIK - Dipanegara Makassar*
SCOPUS ID : 57200209088 *SINTA ID* : 6127241
I Wayan Agus Arimbawa, ST.,M.Eng. - *Universitas Mataram*
SINTA ID : 5973017
Muhammad FauziZulkarnaen, ST.,M.Eng. - *STMIK Lombok*
SINTA ID : 6663733
Yunanri.W, S.T. M. Kom - *UniversitasTeknologi Sumbawa (U.T.S)*
SINTA ID : 6723103
Sitti Aisa, S.Kom.,M.T - *STMIK Dipanegara Makassar*
SINTA ID : 6153893
Sanjaya Pinem, S.Kom, M.Sc . - *Universitas Efarina*
SINTA ID : 6689679
Zamah Sari, S.T., M.T. - *Universitas Muhammadiyah Prof Dr Hamka*
SINTA ID : 6145745
Fredy Windana, S.Kom., MT - *Sekolah Tinggi Teknologi Stikma Internasional*
SINTA ID : 5974460
Hijrah Saputra, ST., M.Sc. - *STMIK Lombok*
SINTA ID : 6667974
Hairul Fahmi, M.Kom. - *STMIK Lombok*
SINTA ID : 5983160
Sofiansyah Fadli, S.Kom.,M.Kom.- *STMIK Lombok*
SINTA ID : 6073057

Editor :
Wire Bagye, S.Kom.,M.Kom- *STMIK Lombok, SINTA ID : 5992010*
Saikin, S.Kom.,M.Kom.- *STMIK Lombok*
Halena Muna Bekata, M.Pd. - *Universitas Tribuana Kalabahi, SINTA ID : 6168815*

Desain Grafis & Web Maintenance
Jihadul Akbar,S.Kom.- *STMIK Lombok*

Secretariat
Ahmad Susan Pardiansyah, M.Kom - *STMIK Lombok*

DAFTAR ISI

1	SISTEM MONITORING TERPADU *SMART BINS* BERBASIS *IoT* MENGGUNAKAN APLIKASI *BLYNK* ***Tatik Juwariyah[1], Luh Krisnawati[2], Sri Sulasminingsih[3]***	**91-99**
2	ANALISIS PERSPEKTIF PADA PENERAPAN E-MONEY MENGGUNAKAN DELONE AND MCLEAN IS SUCCESS MODEL DI BANDARA SULTAN SYARIF KASIM II PEKANBARU ***Fika Felanda Adelia [1], M.Khairul Anam [2], Triyani Arita Fitri [3], Fransiskus Zoromi [4]***	**100-110**
3	ANALISIS SENTIMEN TERHADAP WARGA CHINA SAAT PANDEMI DENGAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY DAN SUPPORT VECTOR MACHINE ***Efid Dwi Agustono[1], Daniel Sianturi[2], Andi Taufik[3], Windu Gata[4]***	**111-119**
4	SISTEM KEHADIRAN MAHASISWA MENGGUNAKAN QR CODE BERBASIS RESTFUL API ***Noor Falih[1], Sarika[2]***	**120-128**
5	IMPLEMENTASI METODE FUZZY TSUKAMOTO UNTUK DETEKSI DINI AUTISME PADA BALITA BERBASIS ANDROID ***Niki Ratama[1], Munawaroh[2]***	**129-139**
6	IMPLEMENTASI RASCH MODEL PADA PERANCANGAN APLIKASI UNTUK PENCARIAN QORI DI WILAYAH KOTA PALEMBANG BERBASIS ANDROID ***M. Rudi Sanjaya[1], Yadi Utama[2], Dedy Kurniawan[3]***	**140-149**
7	IMPLEMENTASI SISTEM KENDALI FUZZY PADA ARAH GERAK ROBOT FINOID ***Almira Nindya Rafi'ah[1], Wahyu S. Pambudi. [2]***	**150-161**
8	PLATFORM WEB SEBAGAI PENAMPIL DATA MONITORING KOTAK SAMPAH BERBASIS IOT ***Dela Citra[1], Irawan Hadi[2], Sarjana[3]***	**162-175**
9	PEMANFAATAN MIKROTIK UNTUK JARINGAN HOTSPOT DENGAN SISTEM VOUCHER PADA DESA UJANMAS KOTA PAGAR ALAM ***Asep Syaputra[1], Dedi Stiadi[2]***	**176-186**
10	KLASIFIKASI KUALITAS UDARA DENGAN METODE SUPPORT VECTOR MACHINE ***Ade Silvia Handayani[1], Sopian Soim[2], Theresia Enim Agusdi[3], Rumiasih[4], Ali Nurdin[5]***	**187-199**

ANALISIS SENTIMEN TERHADAP WARGA CHINA SAAT PANDEMI DENGANALGORITMA*TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY* DAN *SUPPORT VECTOR MACHINE*

Efid Dwi Agustono[1], Daniel Sianturi[2],Andi Taufik[3], Windu Gata[4]

Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta
Jl. Kramat Raya No. 18, Jakarta Pusat, DKI Jakarta, Indonesia
[1]14002333@nusamandiri.ac.id, [2]14002342@nusamandiri.ac.id, [3]andi.iuf@nusamandiri.ac.id,
[4]windu@nusamandiri.ac.id

Abstract

Since the Covid-19 outbreak, virus globally there have been anti-China actions in various countries. The high death rate of the Covid-19 virus incites many countries to take preventive measures that will create some boundaries to the activities of each individual. In Indonesia, the virus has infected as much 34 provinces and 415 cities. Based on the results of the Hootsuite Wearesocial research that released in January 2019, Indonesian social media users have reached around 150 million, or 56percent of Indonesia's population aproximately. Twitter is one of the popular social media where users can post and interact with messages known as "tweets". The tweets can have many likes, dislikes, and contributions on various topics. This research aims to analyze the sentiments of Indonesian citizens towards Chinese citizens in Indonesia on social media Twitter by using the Term Frequency-Inverse Document Frequency and Support Vector Machine Model

Abstrak

Sejak merebaknya virus Covid-19 secara global terjadi aksi anti China di berbagai negara. Tingkat kematian atas virus Covid-19 yang cukup tinggi menyebabkan banyak negara mengambil langkah pencegahan yang membatasi aktivitas setiap individu. Di Indonesia virus tersebut sudah menjangkit 34 provinsi dan 415 kabupaten/kota. Berdasarkan penelitian dari Wearesosial Hootsuite yang dipublikasikan pada Januari 2019 jumlah pengguna media sosial di Indonesia mencapai 150 juta pengguna atau mencapai 56 persen jumlah penduduk Indonesia. *Twitter*merupakan salah satu media sosial populer di mana pengguna dapat membuat status atau disebut "*tweets*". Kicauan tersebut mengandung banyak ekspresi suka, tidak suka, dan kontribusinya pada berbagai topik. Penelitian ini bertujuan untuk mengetahui sentimen warga Indonesia terhadap warga china yang ada di Indonesia dengan permodelan Term Frequuency-Inverse Document Frequency dan Algoritma Support Vector Machine pada media sosial twitter.

Kata Kunci : *covid-19, sentimen, korona, warga china, pandemi.*

1. PENDAHULUAN

Merebaknya pandemi COVID-19 di Indonesia cukup meresahkan semua pihak. Berbagai sector terkena dampak atas merebaknya virus tersebut. Sebagaimana diketahui bahwa virus tersebut pertama kali menjangkit di salah satu propinsi di China yaitu Wuhan[1].Di wuhan telah dilakukan pembatasan keluar masuk wilayah tersebut namun virus sudah menyebar ke seluruh dunia. Penyebabnya karena mobilitas warga china yang cukup tinggi sehingga penyebaran virus menjadi sangat cepat. Di Indonesia virus tersebut sudah menjangkit 34 provinsi dan 415 kabupaten / kota[2]. Pada saat tulisan ini dibuat data pada Worldmeter yang bersumber dari data *World Health Organization* (WHO) sudah tercatat 3.015.411 yang tersebar di 213 negara dengan kematian sebesar 207.968. Di Indonesia kasus positif sebesar 9.096 dengan jumlah kematian sebesar 765 meninggal[3].

Tingkat kematian secara global dari virus tersebut adalah sebesar 6.8%. Namun tingkat kematian di Indonesia lebih tinggi dari tingkat kematian dunia yaitu sebesar 8.4%[2]. Sebelum virus masuk ke Indonesia pemerintah terlihat tidak menyangka akan keganasan dan kecepatan

penyebaran virus tersebut. Hal tersebut dapat dilihat pada kebijakan pemerintah yang kurang tepat dalam mengambil kebijakan. Salah satu kebijakan yang salah adalah memberikan subsidi pada penerbangan domestik yang nilainya mencapai miliaran rupiah. Kebijakan lanjutan justru menghentikan penerbangan domestik dan pelarangan mudik. Pada minggu pertama bulan Mei 2020 penambahan kasus positif diantara 300-400 kasus[2]. Dengan tingkat kesembuhan di angka 50 orang perhari maka menjadi ujian yang berat di bagi paramedis dalam penanganan pasien terkonfirmasi positif[3].

Semakin merebaknya virus Covid-19 membuat pemerintah mengambil keputusan memberlakukan Pembatasan Sosial Berskala Besar (PSBB)[4]. Meskipun tidak menerapkan *lockdown* atau karantina wilayah para pekerja informal sangat terpukul atas pembatasan tersebut, seperti pekerja ojek daring, pekerja harian, dan pedagangyang penghasilannya menurun drastis. Untuk ojek daringtidak diperkenankan mengangkut penumpang. Alhasil tingkat pengangguran semakin meningkat ditambah menjelang bulan ramadhan dipastikan pengeluaran akan meningkat. Perusahaan juga terpukul atas pandemi ini sehingga banyak melakukan PHK. Hal tersebut mengakibatkan banyak pengangguran, dan memunculkan orang miskin baru, sehingga berdampak pada meningkatnya angka kriminalitas.

Berdasarkan penelitian oleh*Wearesosial Hootsuite* yang dipublikasikan pada Januari tahun 2019 jumlah pengguna media sosial di Indonesia mencapai 150 juta atau mencapai 56 persen jumlah penduduk Indonesia[5]. Berikut adalah sepuluh media social yang popular di Indonesia yaitu *Facebook, Twitter, Youtube, Pinterest, Whatsapp, Instagram, linkedin , line, FB Messenger,* dan *wechat*. Pada penelitian kali ini akan menggunakan data dari media sosial *twitter*. Hal tersebut diambil karena data yang begitu terbuka, jumlah *tweet* yang begitu melimpah dan proses pengambilan yang mudah maka diputuskan untuk menggunakan data *twitter* untuk penelitian kali ini.

Media sosial menjadi sangat populer dan menjadi bagian dalam keseharianmasyarakat. Bahkan dalam sehari orang-orang rela menghabiskan waktu berjam-jam dalam mengakses media sosial. Orang-orang dari seluruh dunia membagikan pendapat mereka secara bebas. Banyak pengguna berbagi pemikiran mereka, dan mendiskusikan aspek politik dan ekonomi masyarakat mereka. Di antara media sosial, *Twitter* merupakan platform media sosial yang sangat populer terkait masalah social dan politik [6]. Di sebagian besar media sosial, bahasa yang digunakan oleh pengguna sangat informal [7]. Pengguna media sosial *Twitter* dapat membuat status atau disebut "*tweets*". Kicauan tersebut mengandung banyak ekspresi suka, tidak suka, dan kontribusinya pada berbagai topik[8].

Di berbagai negara bagian Amerika Serikat diantaranya Texas, California,dan New York, banyak orang dari China ditinju, diludahi, ditendang dan bahkan ada yang mengalami penusukan. Warga lokal menganggap pendatang dari China telah mengakibatkan hilangnya penghasilan mereka karena larangan beraktifitas akibat pandemi COVID-19. Beberapa orang china mengalami penyemprotan disinfektan kea rah tubuh mereka dan menerima lontaran kata-kata kotor dari warga lokal[9]. Di Italia orang china juga mengalami hal senada.Di samping air mancur Trevi, Roma, ditempel pengumuman yang isinya melarang pelanggan orang-orang China[10].

Dengan merebaknya COVID-19, ketegangan terhadap orang china mengalami peningkatan. Terdapat seruan di Twitter agar ada fatwa untuk melarang orang China masuk ke Indonesia. Seruanini sebagian adalah pengaktifan kembali sentimen antiChina yang sudah lama berakar di Indonesia. Sentimen antiChina pada media sosial begitu membabi buta, warganet banyak yang menyerukan pengusiran pekerja China untuk keluar dari Indonesia.[11].

2. TINJAUAN PUSTAKA DAN TEORI

2.1 Tinjauan Pustaka

Opinion Mining (OM) dan *Sentiment Analysis* (SA)adalah dua bidang yang muncul yang bertujuan untuk membantu pengguna menemukaninformasi pendapat dan mendeteksi polaritas sentimen. OM dan SA umumnya digunakan secara bergantian untuk mengekspresikan makna yang sama. Namun, beberapa peneliti menyatakan bahwa mereka bertujuan untuk mengatasi dua masalah yang sedikit berbeda [12]. Analisis Sentimen membangun sistem yang mencoba

mengidentifikasi dan mengekstraksi opini dalam teks [8]. Hal ini menyoroti klasifikasi teks dan berkaitan dengan penambangan teks. Menurut T.K. Das biasanya polaritas sentimen diklasifikasikan sebagai positif, kelas negatif atau netral[13].

Penelitian terkait TF-IDF sebelumnya sudah dilakukan oleh Ria, V. Amrizal, Hendra B. Suseno, dan T. Dirjam yang menelitiPenerapan metode TF-IDF *andCosine Similarity*untuk mengetahui syarah sebuah hadits yang diimplementasikan dalam aplikasi berbasis web.

Penelitian lainnya mengenai text mining sudah dilakukan oleh Windu Gata dan Purnomo dalam sebuah jurnal yang menggunakan algoritma KNNuntuk mengetahui akurasi *Text mining*pada Data Berita SMS bulan November 2015. Pada penelitian tersebut mengklasifikasi isi SMS dengan label Ya dan Tidak yang dilakukan secara manual dan dengan algoritma *K-Nearest Neighbour*[14].

Dengan adanya pandemi global ini terjadi aksi anti China di berbagai negara[15]. Sebuah Lembaga Analisis Kebijakan Konflikatau*Institute For Policy Analysis of Conflict* (IPAC), menilai *The*ISISmenjadi aktor penting dalam kegiatan provokasi anti warga keturunan China dan Tionghoa di Indonesia, di tengah-tengah sikap tulus Pemerintah China membantu Indonesia di dalam memerangi meluasnya wabah Penyakit Korona Virus, atau *Corona Virus Disease-19* (Covid-19)[16].Penelitian ini bertujuanuntuk mengetahui sentimen warga Indonesia terhadap warga china yang ada di Indonesia dengan permodelan TF-IDF dan Algoritma*Support Vector Machine*.

Penelitian terkait sentimen analisis sebelumnya sudah dilakukan oleh Imam S, W. Gata, A. B. Paryantiyang meneliti Sentimen Analisis KPU dengan menggunakan *Feature Selection*pada Algoritma *SVM*. Penelitimenggunakan model algoritma *Support Vector Machine* tanpa *feature selection* dibandingkan dengan menggunakan *feature selection Weight by Correlation*yang mana diperoleh hasil *feature selection*dapat meningkatkan nilai dari *Accuracy* dan AUC[17].

Penelitian lain terkait analisis sentimen dilakukan oleh B. Pratama danD.D. Saputra, yang meneliti terkait analisis sentimen terhadap Brimob berdasarkan data pada posting di twitter dengan menggunakan Algoritma *Support Vector Machine* dan *Naïve Bayes*. Dari penelitian tersebut diperoleh kesimpulan bahwa akurasi hasil penelitian sangat ditentukan pada jumlah data *testing* dan data *training*, serta jumlah data set yang digunakan dan komposisi jumlah data positif dan negatif[18].

Penelitian lain dilakukan oleh Sigit Kurniawan, Windu Gata, Dewi Ayu Puspitawati yang meneliti analisis sentimen komentar pada portal berita daring terhadap tokoh politik dengan metode klasifikasi. Pada tahap *preprocessing*Peneliti menggunakan Gata *Framework Textmining*. Dari penelitian tersebut diperoleh hasil jika penggunaan algoritma *SVM* yang dioptimalisasikan dengan PSO memiliki tingkat akurasi lebih tinggijika dibandingkan dengan penggunaan algoritma*Naive Bayes* yang dioptimalisasikan dengan *PSO*[19].

2.2 Support Vector Machine

Support Vector Machine (SVM)adalah salah satu algoritma yang bisa digunakanuntuk permasalahanklasifikasi. Konsep SVM bermula dari klasifikasi dua kelas sehingga membutuhkan*training set* positif dan negatif. SVM berusahamenemukanhyperplane (pemisah) terbaikuntuk memisahkan ke dalam dua kelas danmemaksimalkan margin antara dua kelastersebut.[20]

Formula SVM yang paling sederhana adalah yang linier,dimana *hyperplane* terletak pada ruang data input x. Dalam hal ini ruang hipotesis adalah himpunan bagian dari semua *hyperplanes* berbentuk :

$$f(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x} + b.$$

SVM dalam prakteknya meminimalkan *trade off* antara kesalahan empiris dan kompleksitas ruang hipotesis. Secara formal ini dilakukan dengan memecahkan masalah minimisasi berikut:

SVM Classification =

$$\min_f \|f\|_K^2 + C\sum_{i=1}^{l} |1 - y_i f(\mathbf{x}_i)|_+$$

SVM regression =

$$\min_f \|f\|_K^2 + C\sum_{i=1}^{l} |y_i - f(\mathbf{x}_i)|_\varepsilon$$

C disebut dengan "parameter regularisasi" yang mengontrol *trade off* antara kesalahan empiris

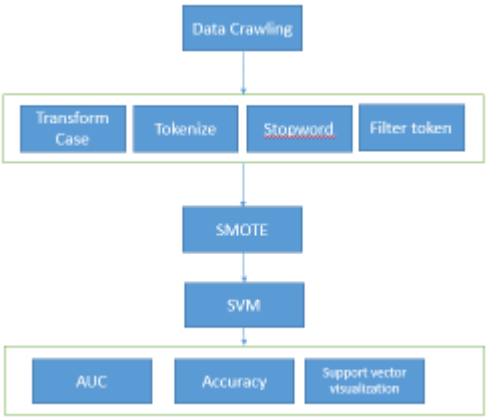
dankompleksitas ruang hipotesis yang digunakan[21].

3. METODOLOGI PENELITIAN

*Text mining*dapat diartikan sebagai sebuah proses dalam mengambil informasi dimana seorang pengguna berhubungan dengan sekumpulan dokumen dan menggunakan perangkatanalisisyangmerupakankomponenpenting dalamdata mining dimana salah satunya adalah kategorisasi[22][23].

3.1 Skema Alur Penelitian

Tahap pertama dalam penelitian ini adalah proses pengumpulan data dan informasi berupa paper atau jurnal penelitian sebelumnya. Data tweet yang diambil adalah *tweet* 2 minggu sebelumnyadengan hanya mengambil *tweet* berbahasa Indonesia. Dari data yang diperoleh tersebut dilakukan *cleansing* data dengan menghilangkan data duplikat. Tahap selanjutnya adalah *text mining*, menghitung bobot pada setiap token dengan menggunakan TF-IDF[14].Model metode penelitian dapat dilihat dalam gambar1, sebagai berikut :



Gambar 1. Metode penelitian sentimen warga china dengan algoritma SVM

3.2 Pengumpulan Data

Data yang digunakan pada penelitian ini bersumber dari posting pada media sosial *twitter*. Proses pengambilan data *twitter* dengan menggunakan kata kunci warga china dengan filter menggunakan bahasa Indonesia dan jumlah *record* yang diambil sebanyak 10.000 baris data. Data *twitter* diambil dengan menggunakan *software* rapidminer versi 9.6. Sebelum diolah kembali di rapidminer data tersebut dibersihkan dulu dengan menghilangkan baris duplikasi. Setelah menghilangkan data duplikat diperoleh data bersih sebanyak 472 baris data.

3.3Pelabelan Data

Proses pelabelan data dilakukan secara manual dengan memberikan nilai positif atau negatif terhadap sebuah cuitan di twitter. Apabila sebuah kalimat mengandung rasa ketidaksukaan pada orang china maka diberikan label negatif dan sebaliknya jika tidak ada penolakan terhadap orang china maka diberikan label positif. Sebagai contoh cuitan "Jangan diskriminasi warga China" diberikan label positif karena tidak ada penolakan dan mengandung sentimen positif. Sebaliknya cuitan berikut "Ini diternate warga china baru mendarat lgi, Kapan mereka dilockdown" diberikan label negatif karena ada penolakan dan rasa kebencian terhadap orang china.

3.4 Text Preprocessing

Text Preprocessing merupakan tahap awal dalam sebuah penelitian terhadap kumpulan banyak teks yang dipersiapkan untuk menjadi sebuah datayang akan diolah lebih lanjut. Suatu teks tidak dapat diproses langsung oleh algoritma pencarian, oleh karena itu dibutuhkan preprocessing text untuk mengubah teks menjadi data numeric[14], [24]. Proses ini terdiri dari beberapa tahap pembersihan dokumen seperti berikut :

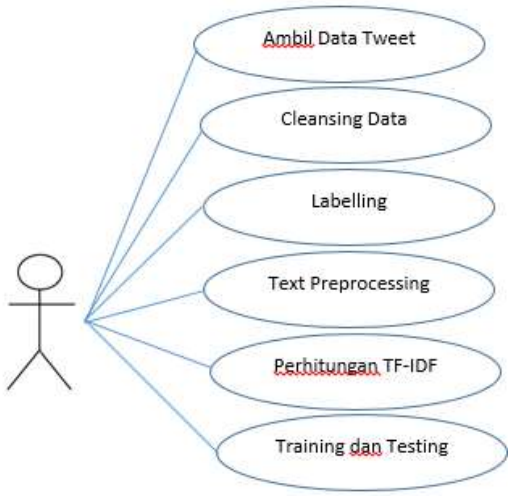
a. *Tokenizing*
*Tokenizing*merupakan sebuah proses mengekstraksi dokumen menjadi kata-kata untuk diolah lebih lanjut. Pada proses inidilakukan dengan cara membuang tanda baca kemudian memisahkan kata-kata tersebut berdasarkan spasi.Pada proses ini juga membuangkaraktertertentusepertihttps://, http://,tanda #, tanda @ yang ada pada *tweet* dan mengubah kata-katake dalam hurufkecil[25].

b. *Stopword Removal* atau *Filtering*
*Stopwords removal*adalah sebuah proses membuang kata-kata yang dianggap tidak penting dan tidak ada keterkaitan dengan penelitian yang dikawatirkan akan mempengaruhi hasil penelitian dengan mengecek kata-kata hasilparsing[22].

c. *Synthetic Minority Over-sampling Technique*SMOTE merupakan sebuah metode dalam penelitian yang biasanya digunakan untuk menangani ketidakseimbangan kelas. Teknik ini mengasosiasi sampel baru dari kelas minoritas agar*dataset*dapat menjadi lebih seimbang dengan cara menyampling ulang contoh kelas minoritas[26]

3.5 Use Case Diagram

Proses pengambilan dan pengolahan data dapat dilihat dalam *use casediagram* seperti pada gambar 2 berikut :



Gambar 2. Use Case Diagram

a. Mengambil data twitter sebanyak 10.000 baris data
b. Melakukan cleansing data dengan membuang baris duplikat
c. Memberikan label negatif dan positif pada setiap baris data
d. Melakukan proses tokenisasi, membuang kata tidak penting, dan melakukan proses SMOTE.
e. Proses pembobotan dengan cara menghitung tingkat kemunculan data
f. Membagi data pelatihan (*training*) dan data uji klasifikasi (*testing*).

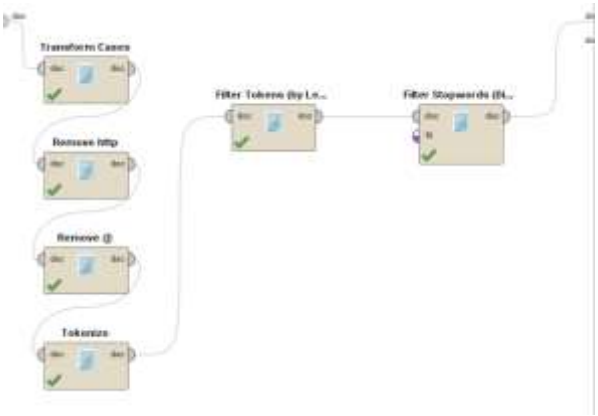
4. HASIL DAN PEMBAHASAN

Data yang digunakan sudah dilakukan proses *cleansing*dengan menghilangkan barisduplikasi. Setelah menghilangkan data duplikat diperoleh data bersih sebanyak 472 baris data.Data yang mengalami proses reduksi ini kan dilakukan pelabelan data "Positif" atau "Negatif". Contoh *tweet* yang diambil dan diberikan label positif dan negatif sebagaimana pada tabel 1 berikut :

Tabel 1. Pemberian label positif dan negatif

Tweet	Label
Ini diternate warga china baru mendarat lgi.. Kapan mereka dilockdown @EnggalPMT @is_pelssy @dr_koko28 @msaid_didu apa ini…	Negatif
RT @firdausmuslim8: Jangan diskriminasi warga China kerana Covid19.Jangan diskriminasi warga China kerana Covid19.	Positif

Data yang sudah diberi label selanjutnya diproses dengan tools rapidminer untuk dilakukan proses *cleansing* data yakni ; merubah semua huruf menjadi huruf kecil (*transform case*), *remove url* (http/https), *remove* karakter '@', *remove* karakter *hashtag* '#', *stemming* data dengan filter *token by length* dengan minimal 4 karakter, dan melakukan filter *stopwords.* Langkah-langkah *text processing*pada aplikasi rapidminer sebagaimana pada gambar 3 berikut :



Gambar 3. Text Processing

Proses pertama merupakan t*ransform case,* yaitu merubah format tulisan menjadi huruf kecil semua. Langkah selanjutnya adalah menghilangkan http, tanda # dan tanda @.Hasilnya sebagaimana tercantum dalam tabel 2di bawah.

Tabel 2. *Text processing*

Sebelum	Sesudah
ya ampun.. kok semakin	ya ampun.. kok semakin

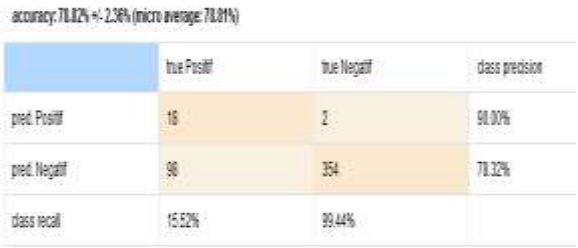
mengerikan... teman2.. minta kpd Allah aga virus ini dimusnahkan. KABAR TERBARU: 43 Ribu Warga China Positif COVID-19 Tanpa Gejala Umum di Wuhan https://t.co/8tuYm1qxHE via @tribunbatamku #covid19 #china	mengerikan... teman2.. minta kpd Allah aga virus ini dimusnahkan. KABAR TERBARU: 43 Ribu Warga China Positif COVID-19 Tanpa Gejala Umum di Wuhan

Kemudian langkah berikutnya menghilangkan kata dengan jumlah dibawah 4 karakter. Selanjutnya adalah memasangfilter *stopword* yaitu mengeliminir kata-kata yang sudah ditentukan sebelumnya.Hasil pada proses ini tercantum sebagaimana pada tabel 3berikut :

Tabel 3.*Filterstopword*

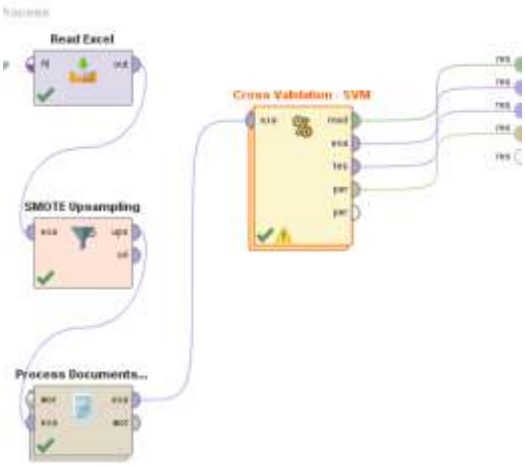
Sebelum	Sesudah
ya ampun.. kok semakin mengerikan... teman2.. minta kpd Allah aga virus ini dimusnahkan. KABAR TERBARU: 43 Ribu Warga China Positif COVID-19 Tanpa Gejala Umum di Wuhan	ampun mengerikan teman allah virus dimusnahkan kabar terbaru ribu warga china positif covid gejala wuhan

Langkah selanjutnya adalah melakukan SMOTE *Upsampling* untuk mengatasi ketidakseimbangan antara jumlah data sentimen negatif dan positif. Sebelum dilakukan SMOTE data *true* positifsebesar 15% dan *true* negatif 99,44% dengan nilai *accuracy* sebesar 78,82%sebagaimana pada gambar4berikut ini.



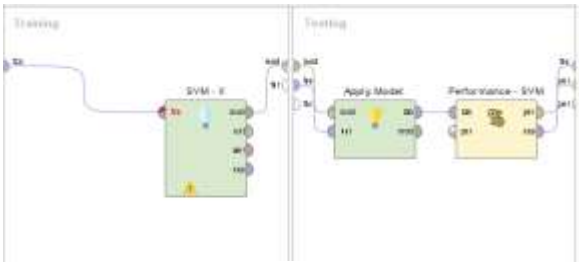
Gambar4. Sebelum *SMOTEUpsampling*

Setelah dilakukan *SMOTEUpsampling* diperoleh nilai *accuracy* yang lebih tinggi yaitu sebesar 91% dengan rincian *true* positif sebesar 89,04% dan *true* negatif sebesar 92,98% seperti disajikan dalam gambar 7.

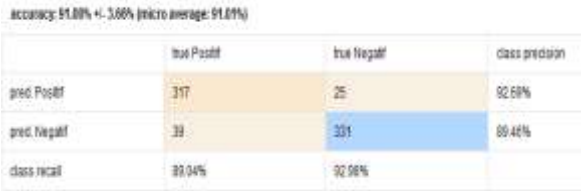


Gambar 5. Desain model Algoritma*SVM*

Sebagaimana digambarkan dalam gambar 5 di atas, data yang telah mengalami proses *text processing*selanjutnya akan dilakukan proses *datacross validation* dengan algoritma SVM untuk mencari pola atau informasi menarik dari sentimenmasyarakat terhadap warga china.



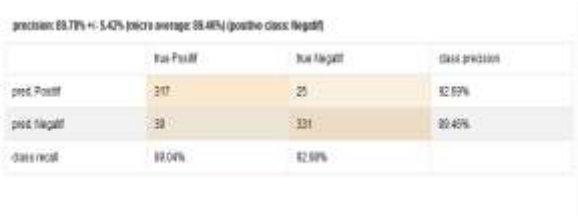
Gambar 6. *Data cross validation*



Gambar 7. Hasil evaluasi akurasi

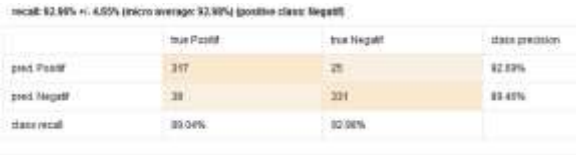
Dari Gambar 7 di atassetelah dilakukan *SMOTEupsampling* dapat diperoleh informasi bahwa total "Prediksi Positif" sebesar 342 data, dimana 317 data dinyatakan *true* positif, dan 25 data dinyatakan *true* negatif, dengan tingkat akurasi sebesar 92,69%. Sementara total data "Prediksi Negatif" adalah sebanyak 370 data, dimana 38 data dinyatakan *true* positif dan 331

data dinyatakan *true* negatif dengan tingkat akurasi sebesar 89,46. Akurasi secara total dapat diperoleh angka sebesar 91,00% dengan plus minus sebesar 3,66%.



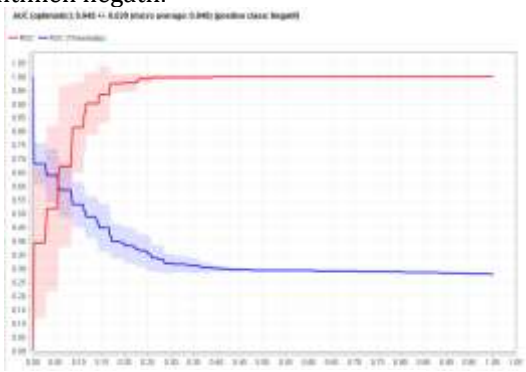
Gambar 8. *Precission*

Untuk mengukur tingkat akurasi perlu ditambahkan hasil *recall* dan *precission*. *Precission*adalahtaraf akurasi antara fakta yang diminta sang pengguna dan jawaban yang didapatkan oleh system. Hasil *precission* dapat dilihat pada Gambar 8Sementara *recall*merupakantaraf keberhasilan sistem untuk menemukan informasisebagaimana pada gambar9 berikut.



Gambar9. *Recall*

Dari data di atas dapat diambil informasi bahwa dari 472 kicauan pengguna media sosial *twitter* yang mengandung kata warga china pada masa pandemi global saat ini mengandung sentimen negatif. *Raw*data yangterkumpuladalah sebanyak 472dataselanjutnya diproses menggunakan metode SVM pada data mining yang menghasilkantingkat akurasi 91,00%dengan rincian 89,04% sentimen positif dan 92,98% sentimen negatif.



Gambar 9.*AUC Process SV*

5. KESIMPULAN DAN SARAN

Pada penelitian ini analisis sentimen memiliki peran penting dalam mengelompokkan tweet yang nantinya akan diproses lebih lanjut menggunakan metode SVM. Dari jumlah tweet yang telah diambil kemudian dilakukan pengolahan data dengan metode *Term Frequency-Inverse Document Frequency* dan algoritma *Support Vector Machine* diperoleh hasil bahwapada masa pandemi ini pengguna twitter di Indonesia memiliki sentimen negatif terhadap warga china.Penggunaan Algoritma *SMOTE Upsampling*dalam sebuah pengolahan data terbukti cukup membantu menciptakan keseimbangan data. Akurasi sebuah hasil penelitian dengan menggunakan *Support Vector Machine* sangat dipengaruhi oleh jumlah data yang digunakan, jumlah data *trainning* dan data *testing*serta komposisi jumlah sentimen positif dan negatif.

6. UCAPAN TERIMA KASIH

Puji syukur alhamdulillah kepada Allah SWT, karena kehendak dan ridhaNya peneliti dapat menyelesaikan penelitian ini. Peneliti juga mengucapkan terima kasih atas bimbingan dan arahan dari Bapak Dr. Windu Gata M.Kom selaku dosen pembimbing mata kuliah Advanced Information Retrieval.

REFERENSI

[1] W. H. Organization, "WHO Timeline - COVID-19," *2020*, 2020. https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 (accessed Apr. 27, 2020).

[2] "Infografis COVID-19 (10 Mei 2020)," *2020*, 2020. https://covid19.go.id/p/berita/infografis-covid-19-10-mei-2020 (accessed May 10, 2020).

[3] "COVID-19 CORONAVIRUS PANDEMIC," *2020*, 2020. https://www.worldometers.info/coronavirus/ (accessed Apr. 27, 2020).

[4] Peraturan Pemerintah (PP) No. 21/2020, *Pembatasan Sosial Berskala Besar (PSBB) Dalam Rangka Percepatan Penanganan COVID-19*. 2020.

[5] "Berapa Pengguna Media Sosial Indonesia?," *2019*, 2019.

https://databoks.katadata.co.id/datapublish/2019/02/08/berapa-pengguna-media-sosial-indonesia (accessed May 04, 2020).

[6] M. Kamyab, R. Tao, M. H. Mohammadi, and A. Rasool, "Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews," *ACM Int. Conf. Proceeding Ser.*, pp. 14–19, 2018, doi: 10.1145/3293663.3293687.

[7] S. Rosenthal, A. Ritter, P. Nakov, and V. Stoyanov, "SemEval-2014 Task 9: Sentiment Analysis in Twitter," 2015, doi: 10.3115/v1/s14-2009.

[8] A. Shelar and C. Y. Huang, "Sentiment analysis of twitter data," *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 1301–1302, 2018, doi: 10.1109/CSCI46756.2018.00252.

[9] M. Lowen, "Covid 19 dan sentimen terhadap orang Asia di Amerika, mereka 'diludahi, dipukul dan dikata-katai' selama pandemi," 2020. https://www.bbc.com/indonesia/majalah-52830126 (accessed Jun. 28, 2020).

[10] M. Lowen, "Virus corona: Italia 'dihantui' virus corona, turis dan warga China menjadi sasaran sentimen rasis," *5 Feb 2020*, 2020. https://www.bbc.com/indonesia/majalah-51383085 (accessed Jun. 28, 2020).

[11] Purnama Ayu Rizky, "Sentimen Anti-China Lebih Seram daripada COVID-19 di Indonesia," 2020. https://www.matamatapolitik.com/sentimen-anti-china-lebih-seram-daripada-covid-19-di-indonesia-in-depth/ (accessed Aug. 28, 2020).

[12] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Comput. Surv.*, vol. 49, no. 2, 2016, doi: 10.1145/2938640.

[13] T. K. Das, D. P. Acharjya, and M. R. Patra, "Opinion mining about a product by analyzing public tweets in Twitter," in *2014 International Conference on Computer Communication and Informatics: Ushering in Technologies of Tomorrow, Today, ICCCI 2014*, 2014, doi: 10.1109/ICCCI.2014.6921727.

[14] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, pp. 1–13, 2017.

[15] Frank Chung, "'Asian dog, you brought corona here': Young women racially abused, spat on in Sydney street," 2020, [Online]. Available: https://www.news.com.au/national/nsw-act/crime/asian-dog-you-brought-corona-here-young-women-racially-abused-spat-on-in-sydney-street/news-story/628e1cbadf94bfdd30fe6dfd4ebe81e8.

[16] T. H. E. Reaction and O. F. Isis, "IPAC Short Briefing No . 1 : COVID-19 AND ISIS IN INDONESIA," no. 1, 2020.

[17] I. Santoso, W. Gata, and A. B. Paryanti, "JURNAL RESTI Penggunaan Feature Selection di Algoritma Support Vector Machine untuk," vol. 1, no. 10, pp. 5–11, 2019.

[18] B. Pratama *et al.*, "Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods," *J. Phys. Conf. Ser.*, vol. 1201, no. 1, 2019, doi: 10.1088/1742-6596/1201/1/012038.

[19] S. Kurniawan, W. Gata, D. A. Puspitawati, N. -, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 176–183, 2019, doi: 10.29207/resti.v3i2.935.

[20] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. March, pp. 1704–1708, 2018.

[21] T. Evgeniou and M. Pontil, "Workshop on Support Vector Machines : Theory and Applications," *Mach. Learn. Its Appl. Adv. Lect.*, no. January 2001, pp. 249–257, 2001, doi: 10.1007/3-540-44673-7.

[22] F. Hermens, "Book Reviews: Book Reviews," *Rev. Soc. Econ.*, vol. 40, no. 1, pp.

72–74, 1982, doi: 10.1080/00346768200000022.

[23] D. Mining and T. Mining, "A Programmers Guide to Data Mining," *Text*, 2017, doi: 10.1109/ISPLC.2012.6201332.

[24] R. Melita *et al.*, "(Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus : Syarah Umdatil Ahkam)," vol. 11, no. 2, 2018.

[25] V. S and J. R, "Text Mining: open Source Tokenization Tools – An Analysis," *Adv. Comput. Intell. An Int. J.*, 2016, doi: 10.5121/acii.2016.3104.

[26] A. Smote and D. A. N. Neighbor, "Klasifikasi Data Tidak Seimbang Menggunakan," vol. 3, no. 1, pp. 44–49.

[27] Lombok, K. I. S. (2015). Perencanaan Strategi Sistem Informasi Pendidikan Pada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Lombok. *Bianglala Informatika*, *3*(2).