

**ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP
PANDEMI COVID-19 MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE BERBASIS
*PARTICLE SWARM OPTIMIZATION***

TESIS

Disusun sebagai salah satu syarat untuk
Memperoleh gelar Magister Komputer
dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh :

GIFTHERA DWILESTARI

NPM : 2018210109



**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

**ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP
PANDEMI COVID-19 MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE BERBASIS
*PARTICLE SWARM OPTIMIZATION***

Oleh:

GIFThERA DWILESTARI

NPM : 2018210109

Bandung, 13 Oktober 2020

Menyetujui,

Dr. Hery Heryanto, S.Kom., M.Kom.
Pembimbing

**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

Dipersembahkan untuk keluarga tercinta

Ayah, Ibu dan Kakak

ABSTRAK

ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP PANDEMI COVID-19 MENGUNAKAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PARTICLE SWARM OPTIMIZATION*

GIFHERA DWILESTARI
NPM : 2018210109

World Health Organization (WHO) telah menyatakan *Corona Virus* sebagai pandemi akibat penyebarannya yang begitu cepat dan menimbulkan banyak korban jiwa di berbagai negara di Dunia. Berbagai informasi dan pemberitaan mengenai Pandemi Covid-19 tidak luput ramai dibicarakan oleh kalangan pengguna *Twitter* di Indonesia. Dengan banyaknya data *tweet* mengenai Pandemi Covid-19 di *Twitter* peneliti pun melakukan analisis sentimen terhadap data tersebut.

Pada penelitian ini peneliti menggunakan algoritme *Support Vector Machine (SVM)* dan *Particle Swarm Optimization (PSO)* digunakan sebagai seleksi fitur untuk meningkatkan nilai akurasi analisis sentimen. Dataset yang digunakan sebanyak 2.919 data *tweet* dengan kata kunci pandemi dan corona. Hasil analisis sentimen pengguna *Twitter* terhadap Pandemi Covid-19 sebanyak 1.250 record menggunakan kamus sastrawi diperoleh yaitu sentimen positif sebanyak 229 *record*, kemudian sentimen negatif sebanyak 763 *record* dan sentimen netral sebanyak 258 *record*.

Klasifikasi terhadap analisis sentimen pengguna *Twitter* mengenai Pandemi Covid-19 menggunakan metode *Support Vector Machine* dengan optimasi *Particle Swarm Optimization* diperoleh hasil akurasi yang menunjukkan rasio prediksi yang benar dengan keseluruhan data sebesar 93,20%. Hasil *class recall true* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan data yang benar negatif sebesar 99,08%. Hasil *class recall true* netral menunjukkan rasio prediksi benar netral dibandingkan dengan keseluruhan data yang benar netral sebesar 84,50%. Hasil *class recall true* positif menunjukkan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif sebesar 83,41%. Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 93,33%. Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 90,08%. Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 96,46%

Dari penelitian tersebut maka diketahui bahwa analisis sentimen mengenai pandemi covid-19 belum akurat karena opini masyarakat belum dapat diekstraksi polanya. Pemasalahan mengenai pandemi covid-19 dapat dikatakan sesuatu yang baru di masyarakat sehingga cara penyampaian komentar dari masyarakat beragam dan sentimen masyarakat mengenai pandemi covid-19 belum bisa diprediksi.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, *Particle Swarm Optimization*.

ABSTRACT

TWITTER SENTIMENT ANALYSIS ON PANDEMIC COVID-19 USING SUPPORT VECTOR MACHINE METHOD AND PARTICLE SWARM OPTIMIZATION

**GIFHERA DWILESTARI
NPM : 2018210109**

The World Health Organization (WHO) has declared the Corona Virus a pandemic due to its rapid spread and causing many casualties in various countries in the world. Various information and news about the Covid-19 Pandemic are not widely discussed by Twitter circles in Indonesia. With the large amount of tweet data about the Covid-19 Pandemic on Twitter, researchers also conducted a sentiment analysis of the data.

In this study, researchers used the Support Vector Machine (SVM) algorithm and Particle Swarm Optimization (PSO) were used as feature selection to increase the value of the sentiment analysis. The dataset used is 2,919 tweet data with the keywords pandemic and corona. The results of the analysis of Twitter users' sentiment towards the Covid-19 Pandemic totaling 1,250 records using a literary dictionary obtained 229 records of positive sentiment, then 763 records of negative sentiment and 258 records of neutral sentiment.

The classification of the sentiment analysis of Twitter users regarding the Covid-19 Pandemic uses the Support Vector Machine method with Particle Swarm Optimization optimization, which shows the correct prediction ratio with total data of 93.20%. The result of class recall is true negative, it shows the ratio of true negative prediction compared to the total data that is true negative is 99.08%. The result of true neutral class recall shows that the prediction is true neutral compared to the total data that is true neutral of 84.50%. The result of class recall is true positive shows that the prediction is true positive compared to the total data that is true positive of 83.41%. Pred precision grade results. negative indicates the ratio of the correct prediction compared to the total result that was predicted to be negative of 93.33%. Pred precision grade results. negative indicates the ratio of the correct prediction compared to the total predicted negative result of 90.08%. Pred precision grade results. negative indicates the ratio of the correct prediction compared to the total result that is predicted to be negative of 96.46%

From this research, it is known that the sentiment analysis regarding the Covid-19 pandemic has not been accurate because public opinion cannot be extracted from the pattern. The problem regarding the Covid-19 pandemic can be said to be something new in society so that the way of submitting comments from the community is diverse and public sentiment regarding the Covid-19 pandemic cannot be predicted.

Keywords: Sentiment Analysis, Support Vector Machine, Particle Swarm Optimization.

KATA PENGANTAR

Puji dan syukur peneliti panjatkan kehadirat Allah SWT, atas limpahan rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan tesis berjudul **Analisis Sentimen Pengguna *Twitter* Terhadap Pandemi Covid-19 Menggunakan Metode *Support Vector Machine* Berbasis *Particle Swarm Optimization***.

Dalam pelaksanaan dan penyusunan penelitian ini tentunya peneliti menghadapi berbagai hambatan dan keterbatasan. Oleh karena itu, pada kesempatan ini peneliti menyampaikan terima kasih kepada berbagai pihak yang tanpa lelah telah membimbing, mendukung dan memotivasi peneliti sehingga peneliti dapat menyelesaikan penelitian dengan baik. Ucapan terimakasih peneliti ingin sampaikan sebesar-besarnya kepada:

1. Bapak Dr. Hery Heryanto, S.Kom., M.Kom, selaku pembimbing yang telah memberikan bimbingan dengan sangat penuh perhatian, ketulusan dan kesabaran.
2. Bapak Dr. H. Budi Permana, S.E., Ak., M.Sc., selaku Ketua STMIK LIKMI Bandung yang telah memberikan arahan dalam penulisan tesis ini.
3. Bapak Dr. Djajasukma Tjahjadi, S.E., M.T., selalu Wakil Ketua STMIK LIKMI Bandung dan juga selaku dosen wali yang telah memberikan motivasi kepada peneliti.
4. Bapak Dr. Dadang Sudrajat, S.Si., M.Kom. selaku Ketua STMIK IKMI Cirebon yang telah memberikan semangat untuk menyelesaikan tesis ini.
5. Seluruh dosen di Program Magister Komputer STMIK LIKMI Bandung yang telah berbagi ilmu pengetahuan kepada peneliti.
6. Seluruh rekan-rekan penulis, baik rekan kampus STMIK LIKMI Bandung maupun rekan kerja di STMIK IKMI Cirebon yang selalu membantu, memberi semangat serta mendukung penulis dalam menyelesaikan tesis.
7. Kedua orang tua, ayah dan ibu, yang tanpa kenal lelah selalu memberikan doa dan dukungan bagi peneliti.
8. Kakak-kakak yang senantiasa memberikan semangat dan doa bagi peneliti.

9. Sahabat-sahabat peneliti yang senantiasa memberikan hiburan bagi peneliti dikala jenuh.

10. Seluruh pihak yang terlibat baik secara langsung maupun tidak langsung yang membantu penulis menyelesaikan tesis ini.

Peneliti menyadari bahwa penyusunan penelitian ini masih banyak kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan sebagai bahan perbaikan di masa yang akan datang. Semoga hasil penelitian ini dapat bermanfaat bagi peneliti khususnya dan dapat digunakan serta dijadikan bahan referensi bagi semua pihak yang berkepentingan.

Bandung, 13 Oktober 2020

Penulis

DAFTAR ISI

ABSTRAK.....	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR LAMPIRAN	vii
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Ruang Lingkup Penelitian	4
1.5 Metode Penelitian.....	5
1.6 Sistematika Penulisan	5
BAB II LANDASAN TEORI.....	7
2.1 <i>Text Mining</i>	7
2.2 Klasifikasi.....	7
2.3 Analisis Sentimen	9
2.4 <i>Support Vector Machine (SVM)</i>	10
2.5 <i>Particle Swarm Optimization (PSO)</i>	11
2.6 <i>RapidMiner</i>	13
2.7 Penelitian Terdahulu.....	16

BAB III OBJEK DAN METODE PENELITIAN	18
3.1 Opini Pandemi Covid-19 pada Media Sosial <i>Twitter</i>	18
3.2 Tahapan Penelitian.....	19
3.2.1 Tahap Pengumpulan Data	20
3.2.2 Tahap <i>Preprocessing</i>	22
3.2.3 Tahap Pembobotan	28
3.2.4 Tahap Klasifikasi dan Optimasi	28
3.2.5 Tahap Validasi dan Evaluasi	29
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Pengumpulan Data.....	30
4.2 <i>Preprocessing</i>	33
4.2.1 <i>Dataset</i>	40
4.2.2 <i>Process Document</i>	42
4.3 Pembobotan	43
4.4 Klasifikasi dan Optimasi	45
4.5 Validasi dan Evaluasi	48
BAB IV KESIMPULAN DAN SARAN	51
5.1 Kesimpulan	51
5.2 Saran	51
DAFTAR PUSTAKA	52

DAFTAR LAMPIRAN

- Lampiran 1 Data Sentimen dengan Kata Kunci Corona
- Lampiran 2 Data Sentimen dengan Kata Kunci Pandemi

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	20
Gambar 3.2 Proses Pengumpulan Data Menggunakan RapidMiner	21
Gambar 3.3 <i>Step 1 of 3 Text-to-Coloumns</i>	23
Gambar 3.4 <i>Step 2 of 3 Text-to-Coloumns</i>	24
Gambar 3.5 <i>Step 3 of 3 Text-to-Coloumns</i>	24
Gambar 3.6 Operator <i>Process Documents</i> Pada RapidMiner.....	28
Gambar 4.1 Hasil <i>Proses Search Twitter dan Write Excel</i>	30
Gambar 4.2 Data Duplikat.....	34
Gambar 4.3 Proses Menghapus Data Duplikasi Menggunakan Microsoft Excel.	34
Gambar 4.4 Proses <i>Read Excel</i> untuk Memasukan <i>Dataset</i>	40
Gambar 4.5 Data Hasil Proses <i>Read Excel</i>	41
Gambar 4.6 Statistik Hasil Proses <i>Read Excel</i>	41
Gambar 4.7 Grafik Hasil Proses <i>Read Excel</i>	42
Gambar 4.8 Model <i>Process Document</i> menggunakan RapidMiner.	42
Gambar 4.9 <i>Process Documents</i>	43
Gambar 4.10 Model Proses	45
Gambar 4.11 Operator <i>Cross Validation</i> Pada Operator <i>Optimize Weights (PSO)</i>	46
Gambar 4.12 Operator yang Digunakan Pada Operator <i>Cross Validation</i>	47
Gambar 4.13 Operator <i>SVM</i> Pada Operator <i>Polynomial to Binominal</i>	47

DAFTAR TABEL

Tabel 2.1 Operator Pengumpulan Data Pada RapidMiner	13
Tabel 2.2 Penelitian Terdahulu	16
Tabel 3.1 Kata Kunci Pencarian Data	18
Tabel 3.2 Hasil Pencarian Data Menggunakan Kata Kunci	18
Tabel 3.3 Proses <i>Cleaning</i>	22
Tabel 3.4 Hasil Analisis Sentimen.	25
Tabel 3.5 Kamus Kata Negatif dan Positif.	25
Tabel 3.6 Contoh Proses Pembobotan.	26
Tabel 3.7 Proses <i>Transform Cases</i>	26
Tabel 3.8 Proses <i>Tokenize</i>	26
Tabel 3.9 Proses <i>Stop Words Removal</i>	27
Tabel 3.10 Proses <i>Stemming</i>	27
Tabel 4.1 Hasil <i>Proses Search Twitter dan Write Excel</i>	31
Tabel 4.2 Hasil Pengumpulan Data dari <i>Twitter</i>	32
Tabel 4.3 Contoh Data yang Dihapus Karena Tidak Mengandung Bahasa Indonesia	35
Tabel 4.4 Data yang Sudah Melewati Proses <i>Cleaning</i>	35
Tabel 4.5 Proses Pelabelan Menggunakan Kamus Sastrawi	36
Tabel 4.6 Hasil Pelabelan Data dengan Kata Kunci Corona	37
Tabel 4.7 Hasil Pelabelan Data dengan Kata Kunci Pandemi	39
Tabel 4.8 Hasil Proses TF dan IDF	44
Tabel 4.9 Hasil Proses TF-IDF	45
Tabel 4.10 Nilai pada parameter operator <i>Optimize Weights (PSO)</i>	45
Tabel 4.11 Nilai Parameter Pada Operator <i>Cross Validation</i>	46
Tabel 4.12 Nilai pada parameter operator <i>SVM</i>	47
Tabel 4.13 Nilai Akurasi	48
Tabel 4.14 <i>Confusion matrix</i> pada akurasi tertinggi	49

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan penggunaan internet di Indonesia semakin tahun kian meningkat dan hal tersebut menyebabkan penggunaan sosial media di kalangan masyarakat Indonesia semakin meningkat pula. Sebuah lembaga survey independen *We Are Social*, dalam waktu satu tahun sejak Januari tahun 2018 hingga Januari 2019 melakukan survey terhadap pengguna internet di Indonesia, hasilnya menunjukkan bahwa pengguna internet di Indonesia mengalami pertumbuhan sebanyak 13% atau sebanyak 17 juta pengguna internet dan pengguna internet yang aktif menggunakan media sosial tumbuh sekitar 15% atau sekitar 20 juta pengguna. Per-Januari 2019 tercatat masyarakat Indonesia yang menggunakan internet telah mencapai 150 juta pengguna dan seluruhnya adalah pengguna internet yang aktif di media sosial (Haryanto, 2019).

Media sosial dinilai dapat dimanfaatkan oleh para penggunanya untuk menuangkan berbagai karya, ide, tanggapan bahkan media untuk mengekspresikan keadaan yang terjadi. Hanya dengan cara membuat akun pribadi, para pengguna media sosial kemudian dapat dengan mudah menuliskan dan mempublikasikan karya maupun tanggapannya pada masyarakat sesama pengguna media sosial. Hal-hal yang membuat media sosial menarik di kalangan masyarakat pengguna internet yaitu diantaranya ketika apa yang dituliskan oleh penggunanya dapat dibaca, dipahami kemudian mendapatkan komentar atau tanggapan dari orang lain. Beberapa media sosial yang memiliki fungsi tersebut diantaranya *Facebook, Twitter, Instagram, Path, Skype, dan YouTube*. Media sosial yang paling populer dan paling diminati oleh masyarakat di Indonesia adalah *Twitter* (Nurhadi, 2017).

Text mining didefinisikan sebagai proses penggalian informasi dimana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan suatu *tools* atau alat analisis. *Text mining* mencari informasi berarti dari sumber- sumber data melalui identifikasi dan eksplorasi pola tertentu. Sumber data *text mining* adalah

kumpulan dokumen dengan pola yang ditemukan pada data teks yang tidak berstruktur. Oleh karenanya *tweet* pada media sosial *Twitter* dapat dijadikan salah satu sumber data berupa teks untuk penelitian *text mining* (Pratama et al., 2017).

Belakangan ini masyarakat dunia sangat diresahkan dengan adanya penyebaran penyakit *Corona Virus* atau Covid-19. *World Health Organization (WHO)* pun telah menyatakan *Corona Virus* sebagai pandemi akibat penyebarannya yang begitu cepat dan menimbulkan banyak korban jiwa di berbagai negara di Dunia. Dilansir dari harian *online Kompas.com*, pada 10 Mei 2020 jumlah kematian yang tercatat dari seluruh dunia akibat virus ini telah mencapai 280.431 jiwa. Sementara pasien yang berhasil dinyatakan sembuh mencapai 1.439.842 jiwa. Di Indonesia sendiri, jumlah kematian akibat virus Covid-19 yang tercatat per-tanggal 9 Mei 2020 sebanyak 959 korban jiwa dari total yang kasus yang tercatat 13.645 jiwa. Dari total tersebut juga ada yang dinyatakan sembuh sebanyak 2.607 jiwa (Aida, 2020).

Asal muasal munculnya Covid-19 ini sampai sekarang belum diketahui, namun virus ini pertama kali terdeteksi pada awal bulan Desember 2019 di Wuhan, China. Virus Corona sebenarnya sudah lama ada namun virus ini hanya ditemukan pada hewan dan tidak dapat menular ke manusia. Kalaupun ada manusia yang terinfeksi oleh virus tersebut, tidak akan sampai menular atau menyebar ke manusia lainnya. Maka hingga saat ini belum diketahui penyebar penyebaran virus ini pada manusia (Yasmin, 2020).

Penyakit Covid-19 memiliki ciri-ciri berupa gejala ringan seperti pilek, sakit tenggorokan, batuk, dan demam. Gejala lain yang lebih serius bagi sebagian orang dan dapat menyebabkan *pneumonia* atau sesak napas. Dalam beberapa kasus, penyakit ini dapat menyebabkan kematian. Orang-orang lanjut usia dan yang memiliki resiko tinggi (seperti asma, diabetes, atau penyakit jantung) lebih rentan untuk mengalami gejala yang berat (Organization, 2020).

Berbagai informasi dan pemberitaan mengenai Pandemi Covid-19 tidak luput ramai dibicarakan oleh kalangan pengguna *Twitter* di Indonesia. Selama beberapa bulan belakangan, pengguna *Twitter* setiap harinya disuguhi dengan segala informasi mengenai perkembangan kasus Pandemi Covid-19 di Indonesia maupun di negara

lainnya. Bahkan di Twitter ada sebuah halaman yang secara *real-time* menampilkan *tweet* dari sumber-sumber informasi resmi dan terpercaya, yang memberikan informasi terkini mengenai situasi kasus Pandemi Covid-19 di Indonesia. Dengan banyaknya data *tweet* mengenai Pandemi Covid-19 di *Twitter* peneliti pun melakukan analisis sentimen terhadap data tersebut.

Penelitian mengenai analisis sentimen telah banyak dilakukan oleh para peneliti lain. Pada penelitian sebelumnya dengan judul “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di *Twitter*” yang dilakukan oleh Hadna, dkk., didapatkan hasil bahwa dari dua metode yang biasa digunakan untuk proses analisis sentimen yaitu metode *Naive Bayes Classifier* dan *Support Vector Machine*, ternyata hasil klasifikasi terbaik didapatkan dengan metode *Support Vector Machine* (Hadna et al., 2016).

Pada penelitian ini peneliti menggunakan algoritme *Support Vector Machine* (SVM) dan *Particle Swarm Optimization* (PSO) digunakan sebagai seleksi fitur untuk meningkatkan nilai akurasi analisis sentimen. Bila dibandingkan dengan metode klasifikasi yang lain, akurasi dari metode *Support Vector Machine* (SVM) lebih besar dan memiliki kecepatan klasifikasi yang tinggi, serta toleransi yang tinggi terhadap atribut yang tidak relevan (Ilmawan & Winarko, 2015). *Particle Swarm Optimization* (PSO) sering digunakan untuk memecahkan problematika optimasi serta masalah seleksi fitur. *Particle Swarm Optimization* (PSO) merupakan suatu teknik optimasi yang sangat sederhana dan mudah untuk menerapkan dan memodifikasi beberapa parameter. Pada teknik *Particle Swarm Optimization* (PSO) ada beberapa cara untuk melakukan optimasi diantaranya adalah dengan cara meningkatkan bobot atribut (*attribute weight*) kepada semua atribut ataupun variabel yang dipakai, menyeleksi atribut (*attribute selection*) dan *feature selection* (Kristiyanti, 2015).

Berdasarkan latar belakang tersebut peneliti ingin mengetahui analisis sentimen pengguna *Twitter* terhadap Pandemi Covid-19 dan menerapkan metode *Support Vector Machine* dalam mengklasifikasikan analisis sentimen dari *tweet* mengenai Pandemi Covid-19 dengan optimasi *Particle Swarm Optimization*. Maka peneliti melakukan

penelitian dengan judul “Analisis Sentimen Pengguna *Twitter* Terhadap Pandemi Covid-19 Menggunakan Metode *Support Vector Machine* Berbasis *Particle Swarm Optimization*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan tersebut, maka permasalahan yang dapat dirumuskan sebagai berikut:

1. Bagaimana analisis sentimen pengguna *Twitter* terhadap Pandemi Covid-19?
2. Bagaimana mengklasifikasikan postingan *Twitter* terhadap Pandemi Covid-19 dengan menggunakan algoritme *Support Vector Machine* berbasis *Particle Swarm Optimization*?

1.3 Tujuan Penelitian

Setelah mengetahui rumusan masalah, maka dapat disimpulkan bahwa tujuan penelitian yaitu:

1. Mengetahui analisis sentimen pengguna *Twitter* terhadap Pandemi Covid-19
2. Menerapkan metode *Support Vector Machine* dalam mengklasifikasikan analisis sentimen dari *tweet* mengenai Pandemi Covid-19 dengan optimasi *Particle Swarm Optimization*.

1.4 Ruang Lingkup Penelitian

Penelitian ini mencakup beberapa pokok bahasan. Berikut adalah ruang lingkup yang dibahas dalam penelitian:

1. *Dataset* mengenai Pandemi Covid-19 diperoleh dari *Twitter*.
2. Pengolahan data menggunakan *tools* RapidMiner versi 9.5 dengan menambahkan *extension Text Processing*

1.5 Metode Penelitian

Dalam penelitian ini penyusunan tesis melakukan beberapa penerapan metode untuk menyelesaikan permasalahan. Metode penelitian yang dilakukan adalah dengan cara:

1. Pengumpulan Data

Pengumpulan data dilakukan berdasarkan *query* atas istilah objek pada aplikasi RapidMiner yang terhubung pada *Twitter API*.

2. *Preprocessing*

Preprocessing data dilakukan dengan tujuan untuk menghilangkan *noise*, menyamakan bentuk kata serta mengurangi volume kata. Tahapan *preprocessing* terdiri dari proses *Cleaning*, *Pelabelan*, *Transform Cases*, *Tokenize*, *Stop Words Removal*, dan *Stemming*.

3. Pembobotan

Tahap pembobotan merupakan proses perubahan teks ke dalam format yang angka yang dapat dimengerti oleh mesin atau algoritme.

4. Klasifikasi dan Optimasi

Proses klasifikasi dalam penelitian ini dilakukan dengan bantuan *software* RapidMiner menggunakan algoritme *Support Vector Machine (SVM)* dan optimasi menggunakan *Particle Swarm Optimization (PSO)* yang digunakan sebagai alat bantu untuk meningkatkan akurasi.

5. Validasi dan Evaluasi

Tahapan yang terakhir adalah melakukan validasi serta evaluasi untuk mengetahui tingkat akurasi dari hasil penelitian.

1.6 Sistematika Penulisan

Sistematika penulisan tesis ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini memuat latar belakang penelitian, rumusan masalah, tujuan penelitian, ruang lingkup penelitian, metode penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Berisi teori-teori yang dijadikan dasar munculnya ide dan acuan untuk mengungkap, menggali dan menyelesaikan masalah tentang penelitian. Teori yang dimuat pada bab ini diantaranya adalah mengenai *Text Mining*, Klasifikasi, Analisis Sentimen, *Support Vector Machine*, *Particle Swarm Optimization*, dan RapidMiner. Selain itu pada bab ini terdapat penelitian-penelitian terdahulu yang memiliki kemiripan dan keterkaitan topik dengan penelitian ini.

BAB III OBJEK DAN METODE PENELITIAN

Berisi tentang objek dalam penelitian yaitu Opini Pandemi Covid-19 pada Media Sosial *Twitter* serta metode penelitian yang akan digunakan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Berisi analisa dari hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan yang didapatkan dari penelitian serta saran untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI

2.1 *Text Mining*

Text mining merupakan suatu proses penambangan data berupa teks yang dilakukan oleh komputer dimana data tersebut dapat memberikan informasi-informasi untuk dilakukan proses analisis (Pudjajana & Manongga, 2018). *Text mining* meliputi kategorisasi, *text clustering*, ekstraksi konsep/entitas, analisis sentimen, *document summarization*, serta *entity-relation modeling*. Sumber data yang digunakan untuk proses *text mining* adalah kumpulan data teks dengan format yang tidak terstruktur atau minimal semi terstruktur. Tujuan dari *text mining* yaitu untuk mendapatkan informasi yang berguna dari sekumpulan dokumen (Gusriani et al., 2016).

Text mining berupaya untuk mengekstrak informasi yang memiliki nilai dan manfaat dari sumber data dengan melalui tahapan identifikasi dan eksplorasi dari suatu pola yang unik dan menarik. Sumber data yang dimaksud dalam *text mining* adalah sekumpulan dokumen dan pola menarik yang tidak ditemukan dalam bentuk *database record*, tetapi ada pada data teks yang tidak terstruktur. Proses *Text Mining* biasanya memerlukan penyusunan teks masukan berdasarkan tata bahasa, yang diikuti dengan menggali pola dari data yang telah terstruktur, evaluasi dan interpretasi hasil. Proses ini digunakan untuk pengklasifikasian, penggerombolan, analisis makna, pengambil kesimpulan dari dokumen dan pemodelan hubungan objek yang berupa kata (Pratama et al., 2017).

2.2 Klasifikasi

Klasifikasi adalah suatu proses yang bertujuan untuk memperkirakan kelas dari suatu objek yang belum diketahui labelnya (Eka Ratnawati et al., 2014). Klasifikasi pada *data mining* adalah penempatan objek kedalam kategori atau kelas yang telah ditetapkan sebelumnya. Klasifikasi sering digunakan untuk memprediksi kelas. Komponen-komponen utama dari proses klasifikasi yaitu:

1. Kelas yaitu variabel tidak bebas hasil dari klasifikasi.

2. Prediktor yaitu variabel bebas, karakteristik atribut data yang diklasifikasi.
3. Training data set yaitu sekumpulan data berisi kelas dan prediktor untuk dilatih agar masuk ke kelas yang tepat.
4. Testing data set yaitu data baru yang akan dikelompokan untuk mengetahui akurasi dari model yang digunakan untuk proses klasifikasi.

Beberapa metode klasifikasi yang populer digunakan dalam penelitian diantaranya yaitu:

1. *Decision Tree*

Decision tree merupakan metode klasifikasi yang terdiri dari *node* keputusan dihubungkan dengan cabang-cabang dan simpul akar sampai ke *node* daun akhir. Pada setiap *node* akan diuji dan setiap hasilnya akan diarahkan ke *node* lain atau ke *node* akhir untuk menghasilkan keputusan (Meilina, 2015).

2. *Naive Bayes*

Metode klasifikasi ini menggunakan perhitungan probabilitas dengan menghitung jumlah frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Metode probabilitas *Naive Bayes* memprediksi peluang di masa mendatang berdasarkan pengalaman pada masa sebelumnya (Saleh, 2015).

3. *Neural Network*

Neural Network adalah algoritme yang meniru fungsi otak manusia yang di dalamnya terdapat ribuan dan jutaan unit pengolahan kecil disebut neuron. Setiap neuron saling terhubung dan memiliki masukan serta keluaran (Rohman & Rochcham, 2019).

4. *K-Nearest Neighbor*

K-Nearest Neighbor atau K-NN merupakan metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi atau jarak dari suatu data dengan data yang lainnya. Nilai K adalah data terdekat dari data uji (Merluarini et al., 2013).

5. *Support Vector Machine*

Metode *Support Vector Machine* ini adalah salah satu metode klasifikasi menggunakan *supervised learning* yang memprediksi kelas berdasarkan pola dari hasil proses *training* (Santoso et al., 2017).

2.3 Analisis Sentimen

Analisis sentimen atau bisa juga disebut *opinion mining* adalah suatu cabang penelitian *text mining* yang mulai populer dilakukan pada tahun 2013. Analisis sentimen merupakan proses memahami, mengekstrak serta mengolah data yang berbentuk teks secara otomatis guna mendapatkan informasi (Indrayuni, 2016).

Opinion mining pada umumnya diperlukan untuk mengetahui sikap seorang pembicara atau penulis berkaitan dengan beberapa topik atau polaritas kontekstual keseluruhan dokumen. Kecendrungan penelitian mengenai analisis sentimen adalah berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negatif. Pendapat mewakili hampir seluruh aktivitas manusia, dikarenakan pendapat dapat memberikan pengaruh atas perilaku dan sikap seseorang. Setiap kali perlu adanya pengambilan keputusan, kita ingin mengetahui pendapat orang lain. Dalam dunia nyata, para pelaku bisnis dan organisasi juga cenderung selalu ingin mengetahui opini publik tentang suatu produk atau jasa. Dengan melakukan analisis sentimen, pelaku bisnis dapat melacak produk-produk, merek dan orang-orang misalnya dan memutuskan apakah opini publik cenderung ke arah positif atau negatif. Hal tersebut mempermudah para pelaku bisnis untuk melacak *deteksi flame (rants* buruk), persepsi produk baru, dan persepsi merek (Praptiwi, 2018).

Analisis sentimen bertujuan menganalisis pendapat, sikap, sentimen, emosi, evaluasi, dan penilaian seseorang dari masyarakat atau para pakar melalui berbagai media yang ada apakah orang tersebut berkenaan dengan suatu topik, produk, individu, organisasi, layanan, ataupun kegiatan tertentu. Tugas utama dalam analisis sentimen yaitu mengelompokkan teks yang terdapat pada sebuah kalimat atau dokumen kemudian menentukan pendapat yang tertuang pada kalimat atau dokumen tersebut apakah sifatnya positif, negatif atau netral (Yislam & Budi, 2016).

Pada dasarnya, analisis sentimen merupakan klasifikasi data tekstual, namun pada kenyataannya analisis sentimen tidak mudah untuk dilakukan dikarenakan adanya perbedaan bahasa yang digunakan untuk menyusun isi dari suatu dokumen. Pada umumnya, teknik yang digunakan untuk melakukan analisis sentimen terbagi menjadi dua teknik yaitu menggunakan teknik simbol dan menggunakan teknik *machine learning*. Pada teknik simbol, analisis dilakukan terhadap setiap kata dari isi dokumen dan dilakukan ekstraksi hubungan untuk mendapatkan sentimen. Pada teknik ini, perlu diketahui setiap makna dan aturan kata dari setiap bahasa yang digunakan untuk menyusun isi dari dokumen karena tiap bahasa memiliki aturan dan makna yang berbeda-beda. Meskipun aturan baku suatu bahasa telah dirumuskan sebelumnya, pada kenyataannya aturan tersebut tidak dapat untuk diterapkan secara sempurna dikarenakan bahasa itu sendiri yang selalu berkembang sedangkan teknik *machine learning* menggunakan statistik dalam melakukan pemrosesan teks. Pada teknik ini, diaplikasikan ilmu atau algoritme yang terdapat pada ilmu statistik seperti teori bayes maupun teori populer lainnya yang dapat diterapkan pada proses *machine learning* (Darma et al., 2018).

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode klasifikasi yang bekerja dengan jalan mencari *hyperplane* dengan margin terbesar. *Hyperplane* merupakan garis batas pemisah data antar-kelas. Margin yaitu jarak antara *hyperplane* dengan data terdekat pada setiap kelas. Adapun data terdekat dengan *hyperplane* pada masing-masing kelas inilah yang dikenal dengan *support vector* (Yusuf & Priambadha, 2013).

Support Vector Machine (SVM) merupakan suatu metode klasifikasi biner yang merupakan algoritme lanjutan pada perbaikan yang dilakukan pihak lain. Algoritme ini bertugas dalam memisahkan *hyperplane* terbaik yang terdiri dari dua kelas yaitu positif dan negatif (Hasanah et al., 2016).

Dalam persamaan ini, parameter W , n , dan b berturut-turut menyatakan bobot vektor (W_1, W_2, \dots, W_n), jumlah atribut, dan scalar.

$$W \cdot X + b = 0 \quad (1)$$

Jika b dianggap sebagai suatu bobot tambahan w_0 , maka persamaan suatu garis pemisah dapat ditulis ulang seperti pada persamaan (2)

$$W_0 + W_1X_1 + W_2X_2 = 0 \quad (2)$$

Sedangkan setiap titik yang terletak di atas garis seperti pada persamaan (3).

Dan sebaliknya, setiap titik yang terletak di bawah garis seperti pada persamaan (4)

$$W_0 + W_1X_1 + W_2X_2 > 0 \quad (3)$$

$$W_0 + W_1X_1 + W_2X_2 < 0 \quad (4)$$

Sehingga didapatkan dua persamaan garis seperti pada persamaan (5) dan (6).

$$H_1 : w_0 + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_0 = +1 \quad (5)$$

$$H_1 : w_0 + w_1x_1 + w_2x_2 \leq 1 \text{ for } y_0 = -1 \quad (6)$$

Berikut ini adalah beberapa fungsi kernel yang umum digunakan diantaranya yaitu :

1. Linier kernel $K(X_i, X_j) = \theta(X_i) \cdot \theta(X_j)$ (7)

2. Polynomial kernel $K(X_i, X_j) = (X_i \cdot X_j + 1)^h$ (8)

3. Gaussian RBF kernel $K(X_i, X_j) = e^{-\frac{\|X_i - X_j\|^2}{2\sigma^2}}$ (9)

4. Sigmod kernel $K(X_i, X_j) = \tanh(kX_i \cdot X_j - \delta)$ (10)

Kernel linier digunakan ketika data yang akan diklasifikasi dapat terpisah dengan sebuah garis. Sedangkan kernel non linier digunakan ketika data hanya dapat dipisahkan dengan garis lengkung atau sebuah bidang pada ruang dimensi tinggi.

2.5 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah metode optimasi yang diperkenalkan oleh Kennedy dan Eberhart di tahun 1995 dengan berdasarkan pada penelitian terhadap perilaku kawanan burung dan ikan. Setiap partikel pada *Particle Swarm Optimization* memiliki kecepatan partikel gerak dalam ruang pencarian dengan kecepatan yang dinamis disesuaikan dengan perilaku mereka. Karena itu, partikel cenderung untuk bergerak menuju ke daerah pencarian yang lebih baik selama proses pencarian (Muhamad et al., 2017).

Particle Swarm Optimization (PSO) merupakan suatu teknik optimasi yang sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam teknik *Particle Swarm Optimization (PSO)* ada beberapa cara untuk melakukan pengoptimasian diantaranya yaitu dengan meningkatkan bobot atribut (*attribute weight*) pada semua atribut atau variabel yang dipakai, menyeleksi atribut (*attribute selection*) dan *feature selection* (Indrayuni, 2016).

Particle Swarm Optimization (PSO) diawali dengan suatu populasi yang terdiri dari sejumlah *term* (yang menyatakan solusi) yang dibangkitkan secara acak. Kemudian melakukan pencarian solusi optimal melalui perbaikan *term* bagi sejumlah kategori tertentu. Pembangkitan posisi (x_i, d) dan kecepatan (v_i, m) dari kumpulan partikel dibangkitkan secara *random* menggunakan batas bawah (X_{min}) dan batas atas (X_{max}). *Update velocity* (kecepatan) untuk semua *term* dari nilai *fitness* dapat diketahui *term* mana yang memiliki nilai global terbaik (*global best*) dan juga dapat diketahui posisi terbaik (*local best*) dari tiap *term* pada semua waktu sekarang dan sebelumnya. Pengulangan akan terus dilakukan sampai kriteria terpenuhi (Cahyono, 2017).

Setiap partikel memiliki posisi dan *velocity* yang dinyatakan dalam vektor. Pencarian solusi pada PSO direpresentasikan dengan perpindahan posisi partikel. Pada setiap iterasi, setiap partikel memperbaharui *velocity* dan posisinya menuju posisi terbaiknya. Pada saat yang sama, terjadi pertukaran informasi posisi terbaik di antara seluruh kumpulan partikel. Pada dasarnya, partikel tidak dapat berpindah secara tiba-tiba, dan bergerak menuju posisi terbaik berdasarkan pengalaman pribadinya maupun pengalaman dari seluruh partikel.

Posisi dan *velocity particle* dinyatakan dalam vektor. Secara matematis, posisi partikel untuk setiap i , diperbaharui melalui persamaan berikut:

$$x_{k+1}^i = x_k^i + v_{k+1}^i \quad (11)$$

Inertia weight (w) dalam perhitungan *velocity*, dapat meningkatkan performa PSO dalam beberapa aplikasi. *Velocity* untuk partikel i pada waktu $k + 1$ diperbaharui dengan perhitungan sebagai berikut:

$$v_{k+1}^i = w_k v_k^i + c_1 r_1 (p_k^i - x_k^i) + c_2 r_2 (p_k^g - x_k^i) \quad (12)$$

$c1$ dan $c2$ merupakan bilangan konstan positif, $r1$ dan $r2$ merupakan bilangan *random* berdistribusi *uniform* antara 0 sampai 1, p_k^i merupakan posisi terbaik dari partikel i pada waktu k , dan p_k^g merupakan posisi terbaik global dari keseluruhan kumpulan partikel pada waktu k . Pada penentuan parameter ω , didapatkan bahwa ω yang bernilai tinggi di awal dan bernilai rendah di akhir memberikan hasil yang lebih baik. Secara matematis, ω pada waktu k dihitung sebagai berikut :

$$w^k = w^{k=1} \times \alpha \quad (13)$$

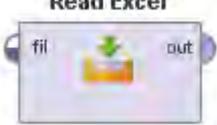
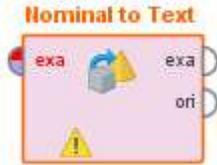
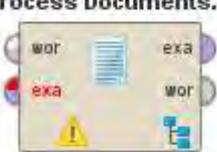
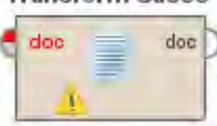
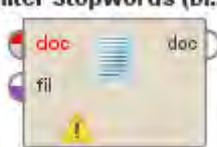
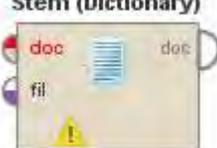
Velocity dibatasi dengan nilai maksimum yaitu v_{max} . Kemampuan eksplorasi partikel dikontrol oleh nilai v_{max} . Penetapan v_{max} sangat mempengaruhi solusi yang diperoleh. Jika v_{max} terlalu kecil atau terlalu besar, partikel tidak dapat mengeksplorasi ruang pencarian dengan baik dan akan mudah terjebak dalam *local optimum*.

2.6 RapidMiner

RapidMiner Studio mengkombinasikan teknologi dan dan pengaplikasikan integrasi yang user-friendly dari teknik-teknik data mining yang telah dimutakhirkan. Proses analisis data menggunakan RapidMiner dilakukan dengan cara mengambil dan memasukan operator, mengatur parameter dan mengkombinasikan operator-operator (RapidMiner, 2012). Operator-operator RapidMiner yang digunakan dalam penelitian ini terdapat pada tabel 2.1.

Tabel 2.1 Operator Pengumpulan Data Pada RapidMiner

OPERATOR	DESKRIPSI
	Operator ini digunakan untuk mencari <i>tweet</i> dari <i>Twitter</i> .
	Operator ini membuat file Excel yang dapat dibaca oleh Excel 95, 97, 2000, XP, 2003 dan versi yang lebih baru

<p>Read Excel</p> 	<p>Operator ini dapat membaca data dari Excel 95, 97, 2000, XP, dan 2003.</p>
<p>Nominal to Text</p> 	<p>Operator ini mengubah jenis atribut nominal yang dipilih ke teks dan juga memetakan semua nilai atribut ke nilai string yang sesuai.</p>
<p>Process Documents...</p> 	<p>Menghasilkan vektor kata dari atribut string.</p>
<p>Transform Cases</p> 	<p>Operator ini mengubah semua karakter dalam dokumen menjadi huruf kecil atau huruf besar.</p>
<p>Tokenize</p> 	<p>Operator ini membagi teks dokumen menjadi urutan token. Ada beberapa opsi cara menentukan titik pemisahan. Salah satunya dengan menggunakan karakter non-huruf yang merupakan pengaturan <i>default</i>. Operator ini akan menghasilkan token yang terdiri dari satu kata tunggal.</p>
<p>Filter Stopwords (Di...</p> 	<p>Operator ini menghapus semua token yang sama dengan stopwords dari file yang diberikan. File harus berisi satu kata kunci per baris.</p>
<p>Stem (Dictionary)</p> 	<p>Mengganti kata atau istilah sesuai dengan aturan pola yang cocok. Mengurangi kata atau istilah ke bentuk dasar menggunakan <i>file</i> eksternal dengan aturan penggantian.</p>

<p>Cross Validation</p> 	<p>Operator ini melakukan validasi silang untuk memperkirakan kinerja statistik dari model pembelajaran</p>
<p>Performance</p> 	<p>Operator ini digunakan untuk evaluasi kinerja. Ini memberikan daftar nilai kriteria kinerja. Kriteria kinerja ini secara otomatis ditentukan agar sesuai dengan jenis tugas pembelajaran.</p>
<p>Optimize Weights (PSO)</p> 	<p>Operator ini melakukan pembobotan fitur dengan pendekatan partikel swarm.</p>
<p>Polynomial by Bino...</p> 	<p>Operator ini membangun model klasifikasi polinomial melalui pembelajar klasifikasi binomial yang diberikan.</p>
<p>SVM</p> 	
<p>Apply Model</p> 	<p>Operator ini menerapkan model pada ExampleSet. Model pertama kali dilatih pada ExampleSet oleh Operator lain, yang biasanya merupakan algoritme pembelajaran. Setelah itu, model ini dapat diterapkan pada ExampleSet lain. Tujuannya adalah untuk mendapatkan prediksi tentang data yang tidak terlihat atau mengubah data dengan menerapkan model preprocessing.</p>

Sumber : Olahan Peneliti

2.7 Penelitian Terdahulu

Tabel 2.2 menunjukkan beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen dan *text mining*:

Tabel 2.2 Penelitian Terdahulu

NO	NAMA PENULIS	JUDUL PENELITIAN	TAHUN	HASIL	PERBEDAAN
1	Veny Amilia Fitri, Rachmadita Andreswari, Muhammad Azani Hasibuan.	<i>Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm.</i>	2019	Hasil yang didapatkan dari penelitian ini menunjukkan bahwa pengguna <i>Twitter</i> lebih banyak memberikan komentar netral. Dengan mengolah data menggunakan metode <i>Naive Bayes</i> pada <i>tools</i> RapidMiner, penelitian ini menghasilkan nilai akurasi sebesar 86.43%. Nilai tersebut lebih tinggi dibandingkan dengan algoritma lainnya, <i>Decision Tree</i> dan <i>Random Forest</i> yang menghasilkan nilai akurasi sebesar 82.91%.	Penelitian ini menganalisis hasil dengan membandingkan tiga metode sedangkan penelitian yang dilakukan sekarang fokus menggunakan satu metode klasifikasi dengan optimalisasi.
2	Dwi Ardiada, Made Sudarma, Dwi Giriantari.	<i>Text Mining</i> pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode <i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i> .	2019	Hasil pengujian dengan metode <i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i> menghasilkan nilai rata-rata <i>precision</i> sebesar 0.4564. Nilai <i>recall</i> sebesar 0.502 dan pada nilai <i>accuracy</i> sebesar 0.8104 sedangkan dari hasil pengujian dengan metode <i>K- Nearest Neighbour</i> nilai rata-rata <i>precision</i> sebesar 0.3421. Nilai <i>recall</i> sebesar 0.4595 dan pada nilai <i>accuracy</i> sebesar 0.797. Maka disimpulkan terdapat peningkatan nilai akurasi jika pengujian dilakukan dengan dua metode.	Penelitian ini menggunakan dua metode untuk melakukan klasifikasi yaitu <i>Support Vector Machine</i> dan <i>K-Nearest Neighbour</i> sedangkan penelitian sekarang menggunakan metode <i>Support Vector Machine</i> untuk klasifikasi dan <i>Particle Swarm Optimization</i> untuk optimasi hasil.
3	Abu Salam, Junta Zeniarja, Rima Septiyan Uswatun Khasanah.	Analisis Sentimen Data Komentar Sosial Media <i>Facebook</i> dengan <i>K-Nearest Neighbor</i> (Studi Kasus Pada Akun Jasa	2018	Hasil yang didapatkan dari metode KNN ini cukup baik. Dari uji coba sebanyak 6 kali didapatkan rata-rata nilai <i>accuracy</i> tertinggi adalah 79.21% sedangkan nilai <i>accuracy</i> terendah adalah 70.3%.	Penelitian ini menggunakan data opini dari media sosial <i>Facebook</i> sedangkan penelitian sekarang menggunakan data opini pada <i>Twitter</i> .

NO	NAMA PENULIS	JUDUL PENELITIAN	TAHUN	HASIL	PERBEDAAN
		Ekspedisi Barang J&T Ekspres Indonesia)			
4	Lilyani Asri Utami.	Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma <i>Support Vector Machine</i> dan <i>KNN</i> Berbasis <i>Particle Swarm Optimization</i> .	2017	Nilai akurasi SVM berbasis PSO lebih tinggi dibandingkan K-NN berbasis PSO. PSO tidak bisa meningkatkan nilai akurasi untuk metode K-NN dalam <i>dataset</i> berbahasa Indonesia.	Pada penelitian ini tidak ada metode analisis sentimen dan tidak ada proses <i>stemming</i> saat <i>preprocessing</i> data.
5	Setyo Budi.	<i>Text Mining</i> Untuk Analisis Sentimen <i>Review</i> Film Menggunakan Algoritma <i>K-Means</i> .	2017	Hasil penelitian menggunakan metode K-Means dengan <i>dataset</i> digunakan sebanyak 300 positif dan 300 negatif hasil akurasi 57.83%, kemudian <i>dataset</i> digunakan sebanyak 700 dokumen positif dan 700 negatif akurasi 56.71%%, dengan <i>dataset</i> digunakan sebanyak 1000 dokumen positif dan 1000 negatif akurasi 50.40%%. Dari hasil pengujian tersebut ditarik kesimpulan bahwa semakin besar <i>dataset</i> yang digunakan semakin rendah akurasi K-Means.	Penelitian ini menggunakan data sentimen <i>review</i> film sedangkan penelitian sekarang menggunakan data opini pada <i>Twitter</i> .
6	Septiyawan Rosetya Wardhana, Diana Purwitasari, Siti Rochimah.	Analisis Sentimen Pada <i>Review</i> Pengguna Aplikasi <i>Mobile</i> Untuk Evaluasi Faktor <i>Usability</i> .	2016	Dari ujicoba menggunakan data <i>review</i> dari 5 aplikasi <i>mobile</i> diperoleh hasil rata-rata nilai akurasi klasifikasi <i>faktor usability</i> sebesar 82% dan akurasi sentimen sebesar 79%.	Penelitian ini menggunakan metode Evaluasi Faktor <i>Usability</i> sedangkan penelitian sekarang menggunakan metode <i>Support Vector Machine</i> untuk klasifikasi dan <i>Particle Swarm Optimization</i> untuk optimasi hasil.

BAB III

OBJEK DAN METODE PENELITIAN

3.1 Opini Pandemi Covid-19 pada Media Sosial *Twitter*

Objek pada penelitian ini adalah *tweet* berbahasa Indonesia yang terdapat di *Twitter*. Dalam menemukan data yang cocok untuk digunakan dalam penelitian ini, peneliti telah melakukan beberapa kali percobaan menarik data dari *Twitter* dengan menggunakan *Twitter API* pada *tools* RapidMiner. Peneliti mencari data pada *Twitter* dengan menggunakan kata kunci yang berkaitan dengan Pandemi Covid-19, data yang dihasilkan dapat dilihat pada tabel 3.1.

Tabel 3.1 Kata Kunci Pencarian Data

PERCOBAAN KE-	KATA KUNCI	JUMLAH DATA YANG DIPEROLEH
1	covid19	1392
2	covid-19	1491
3	pandemi	1499
4	corona	1420
5	coronavirus	1366

Sumber : olahan peneliti

Dari hasil pencarian menggunakan kata kunci didapatkan hasil bahwa data *tweet* terbanyak adalah *tweet* dengan kata kunci “pandemi” dan “corona”. Contoh data yang didapatkan dari kata kunci pandemi dan corona dapat dilihat pada tabel 3.2.

Tabel 3.2 Hasil Pencarian Data Menggunakan Kata Kunci

NO	DATA KEYWORD “pandemi”	DATA KEYWORD “corona”
1	Pemerintah saat ini fokus menangani pandemi Covid-19, tapi agenda strategis yang berdampak besar bagi kehidupan rak... https://t.co/1NvuXsXACi	Virus corona: Ratusan sekolah di Korea Selatan ditutup lagi karena lonjakan kasus. https://t.co/khz6uy8Tx5 https://t.co/xyZZfHFyJF
2	Erick Thohir: Kondisi BUMN di Tengah Pandemi Corona, 90% Sakit https://t.co/9jmCK5QfWc	Lebih dari 200 sekolah di seluruh Korea Selatan ditutup kembali akibat gelombang baru kasus virus corona, sehari-h... https://t.co/uxddZ697rc

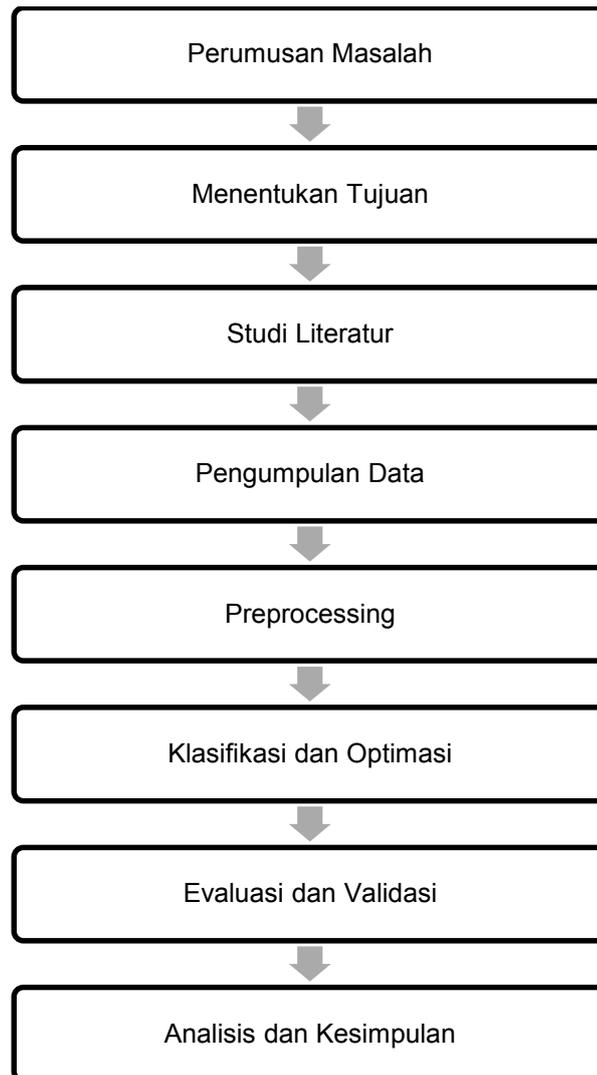
NO	DATA KEYWORD “pandemi”	DATA KEYWORD “corona”
3	Ketika pandemi berdampak kepada penerbit, Mas Agus giat menyemangati kami. Dia sampai meminjamkan mesin tik tua aga... https://t.co/ibJGZqKyXz	Virus corona: Hanya beberapa hari dibuka, ratusan sekolah di Korea Selatan harus ditutup lagi karena lonjakan kasus... https://t.co/6eVoQRJeQc
4	RT @dr_koko28: Tenaga medis sangat diperlukan pada masa pandemi ini. Apalagi di saat kurva masih terus menanjak. Lindungi, jangan dizalimi...	RT @RestyLeseh: Ada untungnya juga Indonesia kena wabah corona . Setidaknya 5 besar penyebab kematian di Indonesia spt Hipertensi , diabet...
5	RT @DRespati2: Pandemi corona hanya melengkapi. Dulu publik sdh ingatkan perlu perbaikan tatakelola BUMN. Manajemen hutang, penegakan prins...	Sekarang tiap bangun tidur dengernya bukan alarm atau ayam berkokok lagi tapi semprotan fogging. DBD ini gamau kalah viral sama Corona
6	RT @HumasBKIPM: Pasar produk hasil perikanan Jawa Tengah di luar negeri tetap menggeliat di masa pandemi. Hal ini ditunjukkan oleh peningka...	RT @GaekParewa1: https://t.co/wKb8WcwevH Semakin kelihatan kualitas kerjanya RS Darurat Corona dibangun di pulau Galang Kota Batam, semen...

Sumber : olahan peneliti

Dari data diatas dapat dilihat bahwa benar bahasa yang ditemukan pada data *tweet* dengan kata kunci “pandemi” dan “corona” adalah bahasa Indonesia. Maka dapat disimpulkan bahwa data-data *tweet* yang dengan kedua kata kunci tersebut sudah cocok dan tepat untuk digunakan dalam penelitian ini. Adapun jumlah data dengan keyword “pandemi” dan “corona” yang telah berhasil didapatkan dan akan digunakan dalam penelitian adalah sejumlah 2.919 data *tweet*.

3.2 Tahapan Penelitian

Penelitian ini melewati beberapa tahapan penelitian seperti yang terdapat pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

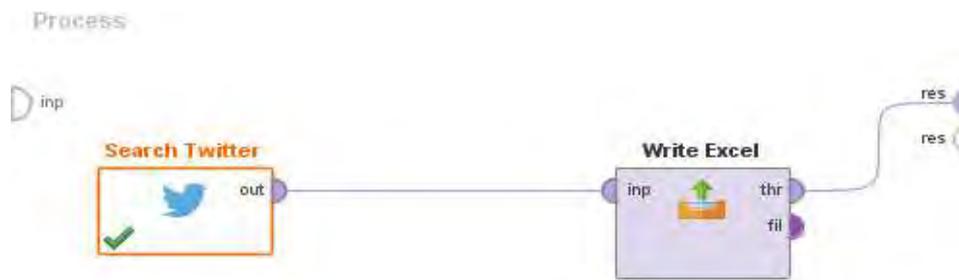
Metode penelitian yang digunakan dalam penelitian diantaranya adalah:

3.2.1 Tahap Pengumpulan Data

Pengumpulan data dilakukan berdasarkan *query* atas istilah objek pada aplikasi RapidMiner yang terhubung pada Twitter API. Hasil *query* berupa *tweet* kotor seperti halnya data pelatihan maupun hasil *query* pengguna mengalami proses *preprocessing* yang sama. Pengumpulan data dilakukan pada tanggal 06 Juni 2020 dengan menentukan parameter pada *search Twitter* bahwa data yang dikumpulkan adalah data *Twitter* sampai dengan tanggal 31 Mei 2020.

Pengumpulan data dilakukan dengan mengambil *tweet* yang memiliki kata kunci terpilih yaitu “pandemi” dan “corona”. Data tersebut didapat dengan cara *crawling* menggunakan *Twitter API* pada *tools* RapidMiner Studio versi 9.5 dan disimpan dalam bentuk Microsoft Excel. Proses ini dapat dilihat pada Gambar 3.2.

Proses pengumpulan data menggunakan *Twitter API* di RapidMiner Studio dimulai dengan membuat proses menggunakan operator *Search Twitter*, kemudian hubungkan dengan akun *Twitter* aktif melalui *login Twitter*. Setelah sudah dipastikan telah terkoneksi maka kita inputkan parameter-parameter yang digunakan dalam untuk mencari data *Twitter* yang diinginkan.



Gambar 3.2 Proses Pengumpulan Data Menggunakan RapidMiner

Parameter dari operator *Search Twitter* yang digunakan dalam penelitian ini yaitu:

1. *Query*: Dalam penelitian ini kata kunci yang akan digunakan untuk mencari data yaitu “pandemi” dan “corona” maka kedua kata kunci itulah yang diinputkan pada *query*.
2. *Result type*: Pada penelitian ini peneliti menggunakan pilihan *default* pada parameter tersebut yaitu *recent or popular* sehingga nantinya data yang didapatkan adalah data *tweet* yang terbaru maupun yang populer dari kedua *query* yang telah ditentukan.
3. *Limit*: Dalam penelitian ini peneliti menentukan untuk mengambil data masing-masing kata kunci sejumlah 1.500 data.
4. *Language*: Dalam penelitian ini peneliti menggunakan *tweet* berbahasa Indonesia maka pada parameter ini peneliti menginputkan “id” yang merupakan kode standar ISO 639-1 untuk Bahasa Indonesia.
5. *Until* : Dalam pengumpulan data ini peneliti tidak menentukan batasan sehingga data yang didapat adalah data paling terbaru ketika operator ini dioperasikan.

Setelah peneliti menentukan parameter untuk operator *Search Twitter*, peneliti membuat lagi proses selanjutnya yaitu proses *Write Excel*. Adapun parameter yang digunakan dalam proses ini adalah parameter default yaitu mengikuti peraturan awal dari RapidMiner.

3.2.2 Tahap *Preprocessing*

Preprocessing data dilakukan dengan tujuan untuk mempersiapkan data mentah sebelum dilakukan proses yang lain dengan cara mengeliminasi data-data yang tidak dibutuhkan atau tidak sesuai dan juga mengubah bentuk data menjadi bentuk yang lebih mudah untuk diproses oleh sistem. Tahapan ini penting untuk dilakukan terutama dalam menganalisis sentimen yang berasal dari media sosial karena kata-kata serta kalimat yang terdapat di media sosial mayoritas merupakan kata-kata dan kalimat yang tidak formal dan juga tidak terstruktur (Mujilawati, 2016). Tahapan *preprocessing* data tekstual dari *Twitter* melibatkan kegiatan berikut ini:

1. *Cleaning*

Pada tahap ini komponen-komponen yang tidak memiliki pengaruh terhadap nilai sentimen dihapuskan. Seperti misalnya @, RT, #, tanda baca, simbol dan *URL* (*Uniform Resource Locator*). Tahap ini dilakukan dengan cara manual menggunakan *Microsoft Excel* dengan menghapus @, RT, #, tanda baca, simbol dan *URL* pada setiap baris data. Selain itu pada proses ini juga dilakukan penghapusan data-data yang duplikat. Contoh dari proses *cleaning* terdapat pada tabel 3.3.

Tabel 3.3 Proses *Cleaning*

SEBELUM	SESUDAH
RT @vivanewscom: Sudah 159 Pasien Positif Corona di DIY Dinyatakan Sembuh https://t.co/opa4ldUriZ #vivanews	vivanewscom Sudah 159 Pasien Positif Corona di DIY Dinyatakan Sembuh vivanews

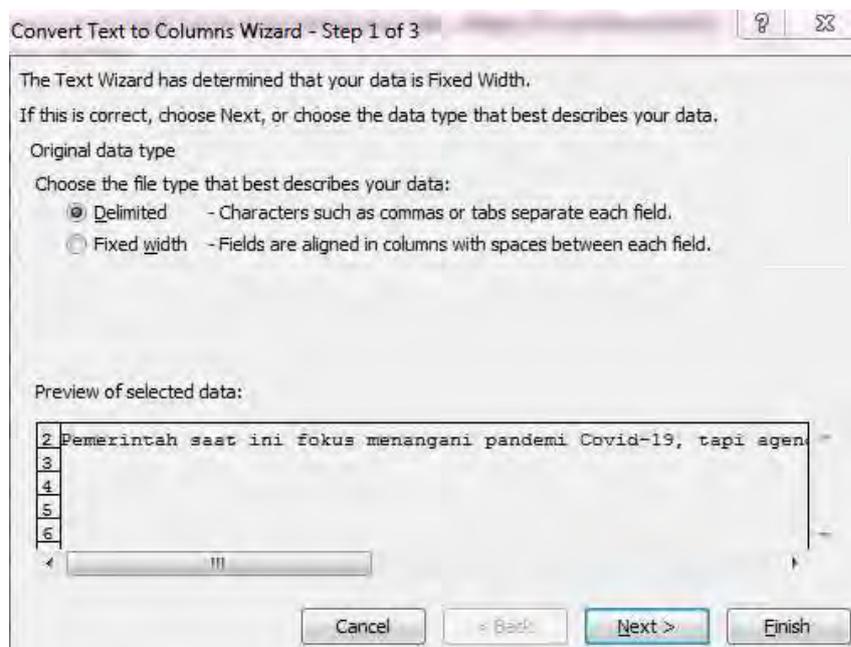
Sumber : Olahan Peneliti

2. Pelabelan

Dalam proses klasifikasi diperlukan dataset yang memiliki label. Proses pemberian label pada dataset akan membutuhkan waktu yang lama terutama jika datanya tidak

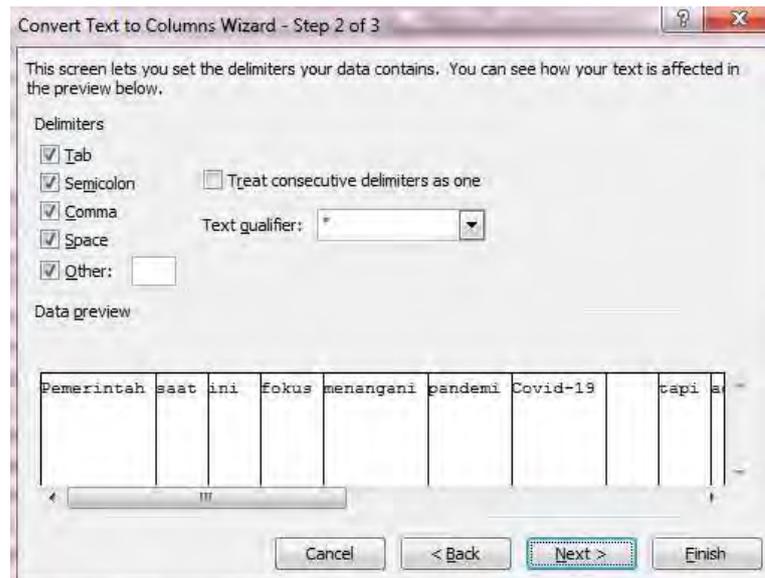
sedikit. Proses pelabelan dapat dilakukan secara manual ataupun menggunakan teknik-teknik *labeling* (Rachmat & Lukito, 2016). Proses pelabelan pada penelitian ini dilakukan dengan bantuan Kamus Sastrawi yang sudah dibuat dalam format *Microsoft Excel*. Dalam menghasilkan sentimen positif, negatif atau netral dilakukan cara-cara berikut ini:

- a. Menyalin teks yang akan dianalisis ke dalam *Microsoft Excel* yang sudah berisikan rumus berdasarkan Kamus Sastrawi.
- b. Melakukan proses *tokenizing* pada teks dengan menggunakan menu *Text-to Columns* pada *Microsoft Excel*.
- c. Memilih proses *delimited* seperti pada Gambar 3.3 untuk kemudian klik *Next >* untuk lanjut ke langkah berikutnya

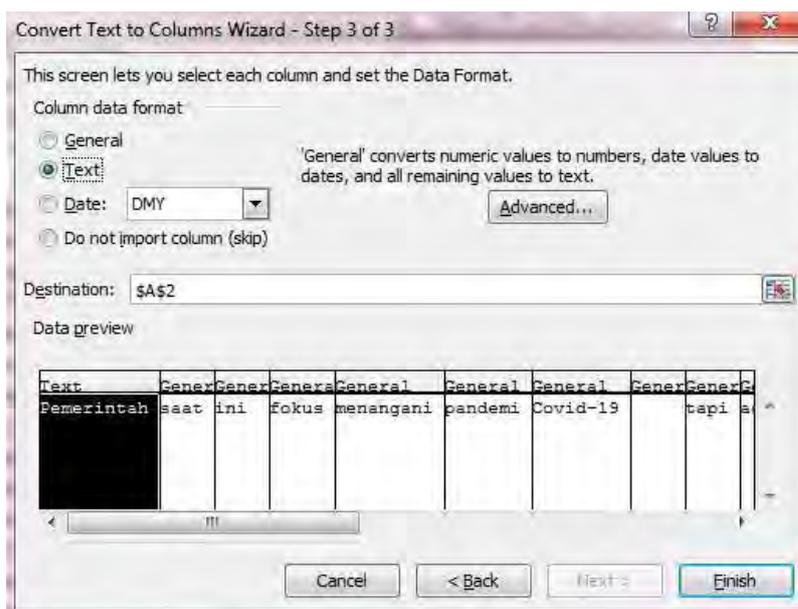


Gambar 3.3 Step 1 of 3 Text-to-Columns

- d. Memilih *delimiters* sesuai dengan kebutuhan. Peneliti memilih semua pilihan *delimiters* seperti pada Gambar 3.4. Kemudian lanjut ke tahap berikutnya sampai dengan *finish* seperti terlihat pada Gambar 3.5.



Gambar 3.4 Step 2 of 3 Text-to-Columns



Gambar 3.5 Step 3 of 3 Text-to-Columns

- e. Setelah kalimat sudah dipecah menjadi token, kemudian akan diketahui hasil sentimen dari kalimat tersebut apakah kalimat tersebut masuk dalam kategori positif, negatif atau netral seperti pada tabel 3.4.

Tabel 3.4 Hasil Analisis Sentimen.

TEXT	POLARITY
Pemerintah saat ini fokus menangani pandemi Covid 19 tapi agenda strategis yang berdampak besar bagi kehidupan rak...	Positif
Erick Thohir Kondisi BUMN di Tengah Pandemi Corona 90 Sakit	Negatif
Ketika pandemi berdampak kepada penerbit Mas Agus giat menyemangati kami Dia sampai meminjamkan mesin tik tua aga...	Netral
DRespati2 Pandemi corona hanya melengkapi Dulu publik sdh ingatkan perlu perbaikan tatakelola BUMN Manajemen hutang penegakan prins...	Netral
HumasBKIPM Pasar produk hasil perikanan Jawa Tengah di luar negeri tetap menggeliat di masa pandemi Hal ini ditunjukkan oleh peningka...	Positif

Sumber: Olahan peneliti.

Di dalam proses menganalisis sentimen ini terjadi juga proses pembobotan, dimana pada satu kalimat apabila ditemukan kata-kata yang terdapat pada Kamus Sastrawi maka dihitung berapa jumlah frekuensi keluarnya kata tersebut. Pada kamus ini sudah dikelompokkan mana kata-kata yang termasuk sentimen negatif dan mana kata-kata yang masuk sentimen positif seperti yang tercantum pada tabel 3.5.

Tabel 3.5 Kamus Kata Negatif dan Positif.

NEGATIF	POSITIF
abnormal	a+
absurd	acungan jempol
acak	adaptif
acak-acakan	adil
acuh	afinitas
acuh tak acuh	afirmasi
adiktif	agilely
adil	agung
agresi	ahli

Sumber : Kamus Sastrawi.

Contoh Proses pembobotan:

Kalimat 1 = Pasien positif corona dinyatakan sembuh

Kalimat 2 = Seorang pasien diduga tertular corona

Contoh dari proses pembobotan bisa dilihat pada tabel 3.6.

Tabel 3.6 Contoh Proses Pembobotan.

No	Pasien	positif	corona	sembuh	seorang	diduga	tertular
1	1	1	1	1	0	0	0
2	1	0	1	0	1	1	1

Sumber : Olahan Peneliti

3. Transform cases.

Transform cases bertujuan mengubah semua huruf menjadi *lowercase* atau huruf kecil. Contoh dari proses *transform cases* bisa dilihat pada tabel 3.7.

Tabel 3.7 Proses Transform Cases

SEBELUM	SESUDAH
vivanewscom Sudah 159 Pasien Positif Corona di DIY Dinyatakan Sembuh vivanews	vivanewscom sudah 159 pasien positif corona di diy dinyatakan sembuh vivanews

Sumber : Olahan Peneliti

4. Tokenize

Tahapan ini mengubah data *tweet* yang berupa kalimat menjadi kata per kata. Contoh dari proses *tokenize* bisa dilihat pada tabel 3.8.

Tabel 3.8 Proses Tokenize

SEBELUM	SESUDAH
vivanewscom sudah 159 pasien positif corona di diy dinyatakan sembuh vivanews	vivanewscom sudah 159 pasien positif corona di diy dinyatakan sembuh vivanews

Sumber : Olahan Peneliti

5. *Stop Words Removal*

Proses ini menghapus atau menghilangkan kata-kata yang tidak penting yang dianggap tidak menggambarkan isi *tweet*. Kata-kata yang dihilangkan dalam proses ini biasanya berupa kata hubung. Contoh proses Stop Word Removal tertuang pada tabel 3.9.

Tabel 3.9 Proses *Stop Words Removal*

SEBELUM	SESUDAH
vivanewscom sudah 159 pasien positif corona di diy dinyatakan sembuh vivanews	sudah pasien positif corona dinyatakan sembuh

Sumber : Olahan Peneliti

6. *Stemming*

Proses ini mengekstraksi kata atau mengubah setiap kata menjadi kata dasar dengan menghilangkan imbuhan (awalan maupun akhiran kata). Contoh proses *Stemming* tertuang pada tabel 3.10.

Tabel 3.10 Proses *Stemming*

SEBELUM	SESUDAH
sudah pasien positif corona dinyatakan sembuh	sudah pasien positif corona nyata sembuh

Sumber : Olahan Peneliti

3.2.3 Tahap Pembobotan

Mesin atau algoritme tidak bisa mengerti kata atau kalimat, maka agar bisa dilakukan proses selanjutnya menggunakan algoritme klasifikasi, data yang berupa teks harus terlebih dahulu diubah kedalam format angka yang dapat dimengerti oleh mesin atau algoritme. Proses pengubahan teks ke dalam format yang angka yang dapat dimengerti oleh mesin atau algoritme ini disebut dengan *feature engineering*. Proses ini sangat penting karena keakuratan hasil algoritme pada dasarnya bergantung kepada pemilihan teknik yang digunakan dalam *feature engineering*.

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk menghitung bobot atau rasio setiap kata yang telah diekstrak (*stemming*). Model pembobotan ini merupakan sebuah pengintegrasian dari model *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. *Term Frequency (TF)* yaitu proses untuk menghitung jumlah kemunculan suatu istilah atau kata (*term*) dalam satu dokumen dan *Inverse Document Frequency (IDF)* digunakan untuk menghitung suatu istilah atau kata (*term*) yang muncul di berbagai dokumen (komentar) yang dianggap sebagai *term* umum, yang dinilai tidak penting (Luqyana et al., 2018).

Proses TF-IDF dilakukan dengan bantuan operator *Process Document* pada RapidMiner seperti pada gambar 3.6. Dimana pada parameter operator *Process Documents* dipilih *create word vector* dan kemudian dipilih pilihan *Term Binary* pada parameter *vector creation*. Hasil dari proses ini adalah jumlah Frekuensi pada setiap *term* yang kemudian diolah kembali sehingga menghasilkan nilai TF-IDF.



Gambar 3.6 Operator *Process Documents* Pada RapidMiner

3.2.4 Tahap Klasifikasi dan Optimasi

Proses klasifikasi dalam penelitian ini dilakukan menggunakan bantuan *software* RapidMiner dengan algoritme *Support Vector Machine (SVM)* dan metode optimasi

Particle Swarm Optimization (PSO) yang digunakan sebagai alat bantu untuk menambah nilai tingkat akurasi.

3.2.5 Tahap Validasi dan Evaluasi

Tahap validasi dan evaluasi dilakukan setelah data selesai diproses dan diketahui berapa banyak sentimen positif, negatif dan netral. Untuk dapat menentukan seberapa besar keakuratan algoritme yang digunakan dalam memproses data tersebut maka perlu diketahui nilai evaluasi akurasi, *precision* dan *recall* untuk menghasilkan konklusi kinerja dari algoritme yang digunakan dalam menganalisis sentimen pada opini pengguna di *Twitter* (Fitri et al., 2019). Presisi adalah presentase dokumen sebenarnya yang diambil relevan dengan *query*. Akurasi adalah nilai perbandingan antara nilai data yang diklasifikasikan secara benar dengan seluruh data seperti pada persamaan. *Recall* adalah presentase dokumen yang relevan dengan *query* dan pada kenyataannya di ambil.

K-fold cross validation adalah salah satu metode yang bisa digunakan untuk mengetahui rata-rata keberhasilan suatu sistem melalui cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. Pengujian dilakukan menggunakan metode *k-fold cross validation* dengan nilai *k* sebanyak 10 *fold*, bertujuan untuk mengetahui akurasi metode *classifier* yang diterapkan pada analisis sentimen jika diuji dengan *data training* dan *data testing* yang berbeda . Penggunaan 10 *fold* ini dianjurkan karena merupakan jumlah *fold* terbaik untuk uji validitas.(Pitria, 2019). Operator-operator RapidMiner yang digunakan pada tahapan ini antara lain operator *Cross Validation* dan operator *Performance*.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data opini pengguna *Twitter* terhadap pandemi Covid-19 didapatkan langsung dari *Twitter* dengan cara *crawling* menggunakan bantuan *tools* RapidMiner. Langkah-langkah yang dilakukan dalam pengambilan data tersebut yaitu:

1. Memasukan operator *Search Twitter* dan Operator *Write Excel* ke Proses RapidMiner.
2. Menghubungkan RapidMiner dengan *Twitter* menggunakan kode Autentifikasi berupa nomor Token yang akan didapatkan dengan melakukan *login Twitter* kemudian memberikan otorisasi akses.
3. Cara untuk bisa dipastikan bahwa RapidMiner telah terhubung dengan *Twitter* klik *Test* di bagian bawah, apabila RapidMiner sudah berhasil terhubung maka akan muncul tanda centang hijau.

Setelah proses *Search Twitter* dan *Write Excel* sudah siap dijalankan maka dilakukan *running* terhadap proses tersebut sehingga muncul hasil seperti pada Gambar 4.1 dan Tabel 4.1 berikut.



Row No.	Id	Created At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1478	1277803854	Jun 30, 2020	Fh	254840301	-	-	en	<a href="http...	Putr
1479	1277803850	Jun 30, 2020	Tapak Id	2232002437	-	-	en	<a href="http...	Jun
1480	1277803828	Jun 30, 2020	kumparan	7596927549	-	-	en	<a href="http...	Pres
1481	1277803827	Jun 30, 2020	kumparan	7596927549	-	-	en	<a href="http...	PSI
1482	1277803825	Jun 30, 2020	NOVA	39331052	-	-	en	<a href="http...	Bay
1483	1277803815	Jun 30, 2020	kumparan	7596927549	-	-	en	<a href="http...	Jaw
1484	1277803821	Jun 30, 2020	RARE	1530181976	-	-	en	<a href="http...	RT @
1485	1277803818	Jun 30, 2020	Rare Tidar B.	85278871	-	-	en	<a href="http...	Pan
1486	1277803817	Jun 30, 2020	Go Suesai	330811360	-	-	en	<a href="http...	Diba
1487	1277803883	Jun 30, 2020	D & P H A	329749531	-	-	en	<a href="http...	Ban
1488	1277803877	Jun 30, 2020	Uji	1260589190	-	-	en	<a href="http...	RT @
1489	1277803822	Jun 30, 2020	Dia	1268224018	Ahmad7_	1242345174	en	<a href="http...	@A
1490	1277803814	Jun 30, 2020	SINDY	78014488	-	-	en	<a href="http...	RT @
1491	1277803793	Jun 30, 2020	karyasmpk	142849198	-	-	en	<a href="http...	RT @

Gambar 4.1 Hasil Proses *Search Twitter* dan *Write Excel*

Tabel 4.1 Hasil *Proses Search Twitter dan Write Excel*

Created-At	From-User		Text		Id
2020-05-29 12:56:15	Joko Widodo		Pemerintah saat ini fokus menangani pandemi Covid-19, tapi agenda strategis yang berdampak besar bagi kehidupan rak... https://t.co/1NvuXsXACi		1266246957971 870000,0
2020-05-30 17:19:33	VIVAcoid		Erick Thohir: Kondisi BUMN di Tengah Pandemi Corona, 90% Sakit https://t.co/9jmCK5QfWc		1266675607514 910000,0
2020-05-30 00:17:44	Fiersa Besari		Ketika pandemi berdampak kepada penerbit, Mas Agus giat menyemangati kami. Dia sampai meminjamkan mesin tik tua aga... https://t.co/ibJGZqKyXz		1266418458503 360000,0
2020-05-31 06:59:58	Achmadya ni.ay70		RT @dr_koko28: Tenaga medis sangat diperlukan pada masa pandemi ini. Apalagi di saat kurva masih terus menanjak. Lindungi, jangan dizalimi...		1266882073500 860000,0
2020-05-31 06:59:57	Nong Nayara		RT @GaekParewa1: https://t.co/wKb8WcwevH Semakin kelihatan kualitas kerjanya RS Darurat Corona dibangun di pulau Galang Kota Batam, semen...		1266882070221 070000,0

Sumber : olahan peneliti

Tabel 4.2 di bawah ini merupakan hasil pencarian menggunakan kelima kata kunci yang berkaitan dengan pandemi covid-19 diantaranya adalah covid19, covid-19, pandemi, *corona* dan *coronavirus*.

Tabel 4.2 Hasil Pengumpulan Data dari *Twitter*

covid19	covid-19	pandemi	corona	coronavirus
<p>Penilaian saya pribadi thd kinerja Pemda dlm mengendalikan #COVID19, scr obyektif tanpa interest politik:</p> <p>Jabar: T... https://t.co/whf2suoqmL</p>	<p>COVID-19: Malaysia dalam fasa pemulihan terakhir - Noor Hisham https://t.co/QmBjkJQs4h #AWANInews #HapusCOVID19... https://t.co/xy9dYIbJqK</p>	<p>Pemerintah saat ini fokus menangani pandemi Covid-19, tapi agenda strategis yang berdampak besar bagi kehidupan rak... https://t.co/1NvuXsXACi</p>	<p>Virus corona: Ratusan sekolah di Korea Selatan ditutup lagi karena lonjakan kasus. https://t.co/khz6uy8Tx5 https://t.co/xyZZfHFyJF</p>	<p>Remember coronavirus?</p>
<p>Alhamdulillah, semua pesakit #Covid19 telah sembuh di Terengganu apabila pesakit terakhir telah discaj dan dibenark... https://t.co/mwYwDtK8CC</p>	<p>Banyak yang bilang dulu waktu awal kemunculan HIV/AIDS juga sempat heboh tapi akhirnya orang juga kembali beraktifi... https://t.co/xHn7FulXJp</p>	<p>Erick Thohir: Kondisi BUMN di Tengah Pandemi Corona, 90% Sakit https://t.co/9jmCK5QfWc</p>	<p>Lebih dari 200 sekolah di seluruh Korea Selatan ditutup kembali akibat gelombang baru kasus virus corona, sehari-h... https://t.co/uxddZ697rc</p>	<p>Dalam tempoh seminggu selepas Aidilfitri, hanya 10 kes babitkan rakyat Malaysia positif jangkitan COVID-19... https://t.co/riHhluwbgI</p>
<p>#COVID19 ???? Kluster baharu. Kluster syarikat pembersihan. 24 orang telah disahkan positif. Semuanya bukan warganega... https://t.co/0TPcqP5bls</p>	<p>Pemerintah saat ini fokus menangani pandemi Covid-19, tapi agenda strategis yang berdampak besar bagi kehidupan rak... https://t.co/1NvuXsXACi</p>	<p>Ketika pandemi berdampak kepada penerbit, Mas Agus giat menyemangati kami. Dia sampai meminjamkan mesin tik tua aga... https://t.co/ibJGZqKyXz</p>	<p>Virus corona: Hanya beberapa hari dibuka, ratusan sekolah di Korea Selatan harus ditutup lagi karena lonjakan kasus... https://t.co/6eVoQRJeQc</p>	<p>#Coronavirus merupakan salah satu alasan yang tepat untuk berhenti menggunakan tembakau. Mengapa? Karena dampak... https://t.co/t2AKfxHqOQ</p>
<p>RT @KKMPutrajaya: Selamat Menyambut Pesta Kaamatan dan Hari Gawai.</p> <p>Kotobian Tadau Tagazo Do Kaamatan & Gayu Guru</p>	<p>RT @OssyDerman: Jika ada Diskusi Publik berjudul "Kehebatan Indonesia Memberlakukan New Normal Saat Penyebaran</p>	<p>RT @dr_koko28: Tenaga medis sangat diperlukan pada masa pandemi ini. Apalagi di saat kurva masih terus menanjak.</p>	<p>RT @RestyLeseh: Ada untungnya juga Indonesia kena wabah corona . Setidaknya 5 besar penyebab kematian di Indonesia spt</p>	<p>RT @michaeljknwles: Remember coronavirus?</p>

covid19	covid-19	pandemi	corona	coronavirus
Gerai Nyamai. Sambutlah...	Covid-19 Masih Tinggi”, ap...	Lindungi, jangan dizalimi...	Hipertensi , diabet...	
RT @KKMPutrajaya: Kotobian Tadau Tagazo Do Kaamatan & Gayu Guru Gerai Nyamai. Sambutlah dengan norma- norma baharu bagi mencegah penularan...	RT @emilia_ts: Halo! Aku buka kelas gratis buat temen2 yg kehilangan pekerjaan atau sulit cari kerja krn COVID-19. Jadwal kelas dan cara...	RT @DRespati2: Pandemi corona hanya melengkapi. Dulu publik sdh ingatkan perlu perbaikan perlu tatakelola BUMN. Manajemen hutang, penegakan prins...	Sekarang tiap bangun tidur dengernya bukan alarm atau ayam berkokok lagi tapi semprotan fogging. DBD ini gamau kalah viral sama Corona	RT @michaeljkn wles: Remember coronavirus?

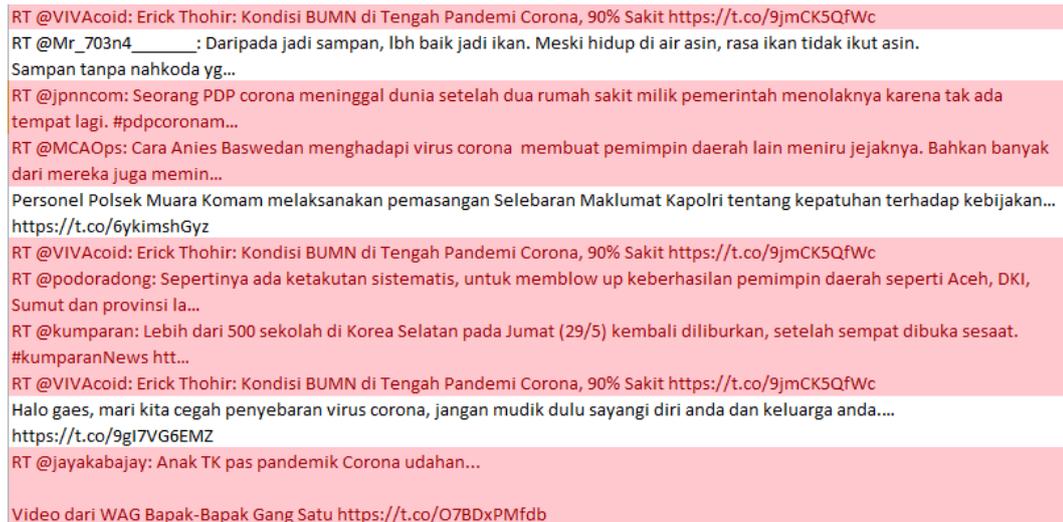
Sumber : olahan peneliti

4.2 Preprocessing

1. Cleaning

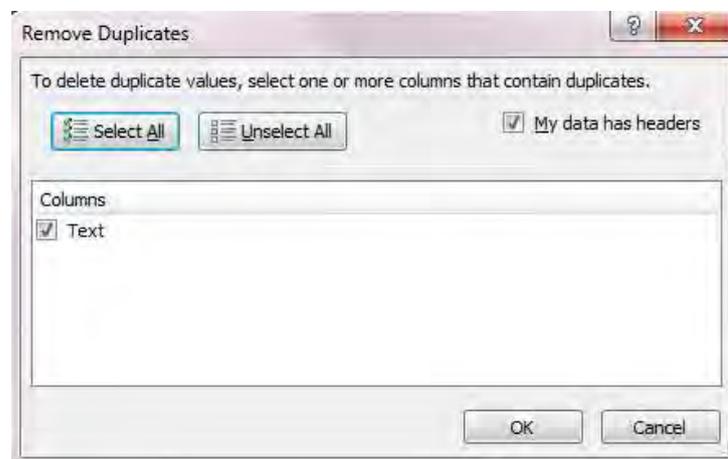
Pada tahapan ini yang pertama dilakukan adalah melakukan pengecekan duplikasi terhadap data yang telah dihimpun. Apabila ditemukan data duplikasi maka data tersebut dihapus. Kemudian dilakukan juga penghapusan terhadap RT, URL, @, tanda baca ataupun simbol lainnya karena tidak berpengaruh terhadap sentimen. Proses proses *cleaning* dilakukan secara manual menggunakan bantuan Microsoft Excel.

Cara pengecekan data duplikasi dengan melalui beberapa tahapan, yaitu: Pertama pilih *Conditional formating* kemudian *Highlight Cells Rules* kemudian pilih *Duplicate Values* lalu klik *OK*. Melalui tahapan tersebut maka akan terlihat hasil seperti pada Gambar 4.2. Baris yang berwarna merah pada gambar menunjukkan bahwa data pada baris tersebut memiliki kesamaan dengan data pada baris yang lainnya.



Gambar 4.2 Data Duplikat

Setelah diketahui data-data yang duplikat maka data tersebut perlu dihapuskan. Maka proses selanjutnya adalah menghapus data yang duplikat menggunakan menu *Remove Duplicates* pada Microsoft Excel. Proses ini secara otomatis menghapus data-data duplikasi dan menyisakan satu data unik untuk disimpan. Proses *Remove Duplicates* dapat dilihat pada Gambar 4.3.



Gambar 4.3 Proses Menghapus Data Duplikasi Menggunakan Microsoft Excel.

Setelah penghapusan data duplikat, tahap selanjutnya yaitu menghapus RT, @ serta tanda baca dan simbol lainnya yang ditemukan pada data. Proses ini dilakukan secara manual menggunakan *Find and Replace* pada Microsoft Excel. Simbol-simbol dan tanda baca yang dihapus dalam proses ini diantaranya yaitu bintang (*), garis

miring (/), tanda petik(' dan "), garis bawah (_), persen (%), tanda pagar (#), tanda seru (!), tanda tanya(?), titik dua (:), titik(.), koma(,), sama dengan (=), kurung siku([]) kurung buka ((), kurung tutup ()).

Setelah dilakukan penghapusan terhadap simbol dan tanda baca, kemudian dilakukan penghapusan *URL* yang termuat pada data. Proses penghapusan *URL* ini juga dilakukan secara manual pada Microsoft Excel. Selain itu juga dilakukan pengecekan terhadap data yang memuat bahasa asing. Apabila ditemukan data yang tidak mengandung Bahasa Indonesia maka data tersebut dihapuskan karena tidak sesuai dengan kriteria data yang dibutuhkan untuk penelitian. Contoh kalimat yang tidak mengandung bahasa Indonesia dapat dilihat pada Tabel 4.3.

Tabel 4.3 Contoh Data yang Dihapus Karena Tidak Mengandung Bahasa Indonesia

NO	TEKS
1	vijayrupanibjp sir hu Gujarat University no student chuane final year ma chuhamari exam 25 June thi chalu t... https://tco/RPjpo6XRy8
2	न्यायकरोमामा Desh me corona k cases bahut bad rahe h or ham exam dene Gaye or koi corena patient aa gya toh Kya karegi sarkar
3	ABPNNews rsprasad Zimmedari se palla hataya bola ganti bajaocorona bhagaowow master strokes
4	EngnrDan Watu wa Mlango Kubwa Mathare yenye inaongoza na visa vya Corona wanasema hawajaona hata mgonjwa mmoja ??
5	kigogo2014 Watanzania tujilinde wenyewe tuliowapa dhamana wanatusakizia Kwenye noma huku wao wakijilinda Sisi tu... https://tco/BQI46qvydP

Sumber : olahan peneliti

Setelah dilakukan penghapusan data yang tidak sesuai kriteria maka telah didapatkan data yang sudah bersih sebanyak 677 data bersih dengan kata kunci corona dan 573 data bersih dengan kata kunci pandemi. Maka total data yang akan diproses ke tahap selanjutnya adalah 1.250 data. Data yang sudah selesai melewati proses *cleaning* tertuang pada tabel 4.4.

Tabel 4.4 Data yang Sudah Melewati Proses *Cleaning*

id	TEXT
1	Pemerintah saat ini fokus menangani pandemi Covid19 tapi agenda strategis yang berdampak besar bagi kehidupan rak
2	Erick Thohir Kondisi BUMN di Tengah Pandemi Corona 90 Sakit
3	Ketika pandemi berdampak kepada penerbit Mas Agus giat menyemangati kami Dia sampai meminjamkan mesin tik tua aga
4	drkoko28 Tenaga medis sangat diperlukan pada masa pandemi ini Apalagi di saat kurva masih terus menanjak Lindungi jangan dizalimi

id	TEXT
5	GaekParewa1 Semakin kelihatan kualitas kerjanya RS Darurat Corona dibangun di pulau Galang Kota Batam semen
6	DRespati2 Pandemi corona hanya melengkapinya Dulu publik sdh ingatkan perlu perbaikan tatakelola BUMN Manajemen hutang penegakan prins
7	HumasBKIPM Pasar produk hasil perikanan Jawa Tengah di luar negeri tetap menggeliat di masa pandemi Hal ini ditunjukkan oleh peningka
8	ntsana bagian paling sedih karna beberapa di antara kita punya janji untuk habis pandemi ketemu ya ternyata sebelum pandemi ini
9	RadioElshinta Agar tetap fit walaupun sedang DirumahAja olahraga itu penting untuk meningkatkan imun tubuh kita khususnya di masa
10	yunggichi makan kak ntar sakit jaga kesehatan kak di tengah pandemi ini
1250	Semakin lama musim pandemi Semakin banyak orang yang tidak peduli Dan akhirnya terjadi Beberapa orang klarifikasi

Sumber : olahan peneliti

2. Pelabelan

Data yang telah melewati proses *cleaning* kemudian diberi label positif, negatif atau netral. Pada tahapan ini peneliti dibantu oleh kamus sastrawi untuk menentukan sentimen positif, negatif dan netral. Cara menentukan sentimen menggunakan Kamus Sastrawi yaitu pertama menyalin data yang sudah melewati tahap *cleaning* ke *Microsoft Excel* yang sudah diseting dengan kamus sastrawi. Kedua, lakukan proses *stemming* dengan menu *Data* pada *Microsoft Excel* setelah itu lewati tahapan *Text-to-Column* > *Delimitate* > pilih semua opsi *Delimiters* > Pilih *Collumn Data Format* > *Finish*. Tabel 4.5 menunjukkan proses pelabelan menggunakan Kamus Sastrawi.

Tabel 4.5 Proses Pelabelan Menggunakan Kamus Sastrawi

id	TEXT	TOTAL NEGATIF	TOTAL POSITIF	SENTIMEN	POLARITY
251	maspiyuuu Konspirasi Corona Zionis Yahudi Dalangnya? Untuk Wujudkan The New World Order NEW NORMAL	0,142	0	-0,143	Negatif
252	Tanggal 1 juni the new normal akan diberlakukan jangan disalah artikan ini semua upaya kita melawan corona	0	0	0	Netral
253	Bolalob Negara Lain Masih Sibuk Melawan Corona Vietnam Sudah Fokus untuk Piala AFF 2020	0	0,076	0,077	Positif

id	TEXT	TOTAL NEGATIF	TOTAL POSITIF	SENTIMEN	POLARITY
254	labubelah Semoga corona cepet selesai supaya semuanya dapat berjalan seperti sedia kala Dan Bismillah semoga aku	0	0,125	0,125	Positif
255	polsekbae protokol kesehatan wajib dilakukanutk cegah corona ya Gaes	0	0,111	0,111	Positif
256	Alasan gw ga takut corona Buat yg mata hatinya masih terbuka bantu reetwet	0,153	0	-0,154	Negatif
257	Daripada ke pasar dapat corona lebih baik belanja sayur online	0	0,1	0,1	Positif
258	IBUFoundation Selamat Hari Lanjut Usia atau Hari Lansia Nasional Di tengah situasi pandemi Virus Corona peringatan Hari Lansia menja	0,105	0,105	0	Netral
259	1 Pegawai Supermarket di Ciamis Terkonfirmasi Positif Corona	0	0,125	0,125	Positif
260	Dampak Corona Renault PHK 15000 Karyawan di Seluruh Dunia Berita Trans	0,1	0	-0,1	Negatif
261	NUgarislucu Doa saat Corona yang berakhir bencana	0,285	0	-0,286	Negatif

Sumber : olahan peneliti

Karena pada Kamus Sastrawi tersebut belum memiliki kata-kata baru seperti covid, covid-19, corona, pandemi, dan lain sebagainya. Hasil dari proses pelabelan menggunakan kamus sastrawi untuk data dengan kata kunci corona sejumlah 677 data, dihasilkan 253 data label negatif, 307 data label netral dan 117 data label positif. Tabel 4.6 menunjukkan hasil pelabelan menggunakan Kamus Sastrawi dari data dengan kata kunci Corona.

Tabel 4.6 Hasil Pelabelan Data dengan Kata Kunci Corona

id	TEXT	POLARITY
1	Virus corona Ratusan sekolah di Korea Selatan ditutup lagi karena lonjakan kasus	Netral
2	Lebih dari 200 sekolah di seluruh Korea Selatan ditutup kembali akibat gelombang baru kasus virus corona berharih	Netral
3	Virus corona Hanya beberapa hari dibuka ratusan sekolah di Korea Selatan harus ditutup lagi karena lonjakan kasus	Netral
4	RestyLeseh Ada untungnya juga Indonesia kena wabah corona Setidaknya 5 besar penyebab kematian di Ibdonesia spt	Netral

id	TEXT	POLARITY
	Hipertensi diabet	
5	Sekarang tiap bangun tidur dengernya bukan alarm atau ayam berkokok lagi tapi semprotan fogging DBD ini gamau kalah viral sama Corona	Netral
6	GaekParewa1 Semakin kelihatan kualitas kerjanya RS Darurat Corona dibangun di pulau Galang Kota Batam semen	Netral
7	DRespati2 Pandemi corona hanya melengkap Dulu publik sdh ingatkan perlu perbaikan tatakelola BUMN Manajemen hutang penegakan prins	Netral
8	detikcom Uji spesimen yang dilakukan pemerintah hari ini melebihi target yakni sebanyak 11361 spesimen sehingga total yang sudah dip	Netral
9	gosuselcom 30 ODP 5 Petugas RSUD Daya Positif Corona Layanan UGD Disetop Sementara	Netral
10	VIVAcoid Erick Thohir Kondisi BUMN di Tengah Pandemi Corona 90 Sakit	Netral
11	vivanewscom Sudah 159 Pasien Positif Corona di DIY Dinyatakan Sembuh	Netral
12	emmambemz kazlialexandria jansenjsp Dulu sby buat pengaduan via sms nulis apapun gak ada yang di tangkap sek	Netral
13	jayakabajay Anak TK pas pandemik Corona udahan Video dari WAG BapakBapak Gang Satu	Netral
14	Ediyuhono Ga beli di subsidi pemerintah Belut termasuk korban corona	Netral
15	HusniTanjung demoSoCRAZY Nyali 011 micron lbh kecil dr virus corona	Netral
16	Wali Kota Sebut Kota Bogor Saat Ini Zona Kuning Penyebaran Corona COVID19	Netral
17	Sudah 159 Pasien Positif Corona di DIY Dinyatakan Sembuh	Netral
18	Pagi ini di tukang surabi orangorang lagi pada nunggu tibatiba ku batuk trs pada ngeliatin Ya Allah corona cepat	Netral
19	Mr703n4 Daripada jadi sampan lbh baik jadi ikan Meski hidup di air asin rasa ikan tidak ikut asin Sampan tanpa nahkoda yg	Netral
20	jpnncm Seorang PDP corona meninggal dunia setelah dua rumah sakit milik pemerintah menolaknya karena tak ada tempat lagi pdpcoronam	Netral
677	Semakin lama musim pandemi Semakin banyak orang yang tidak peduli Dan akhirnya terjadi Beberapa orang klarifikasi	Netral

Sumber: olahan peneliti

Hasil dari pelabelan data dengan kata kunci pandemi dengan jumlah data 573 didapatkan 181 data label negatif, 220 data label netral dan 172 data label positif. Tabel 4.7 menunjukkan hasil pelabelan menggunakan Kamus Sastrawi dari data dengan kata kunci Pandemi.

Tabel 4.7 Hasil Pelabelan Data dengan Kata Kunci Pandemi

id	TEXT	POLARITY
1	Pemerintah saat ini fokus menangani pandemi Covid19 tapi agenda strategis yang berdampak besar bagi kehidupan rak	Positif
2	Erick Thohir Kondisi BUMN di Tengah Pandemi Corona 90 Sakit	Negatif
3	Ketika pandemi berdampak kepada penerbit Mas Agus giat menyemangati kami Dia sampai meminjamkan mesin tik tua aga	Netral
4	drkoko28 Tenaga medis sangat diperlukan pada masa pandemi ini Apalagi di saat kurva masih terus menanjak Lindungi jangan dizalimi	Netral
5	GaekParewa1 Semakin kelihatan kualitas kerjanya RS Darurat Corona dibangun di pulau Galang Kota Batam semen	Netral
6	DRespati2 Pandemi corona hanya melengkapi Dulu publik sdh ingatkan perlu perbaikan tatakelola BUMN Manajemen hutang penegakan prins	Netral
7	HumasBKIPM Pasar produk hasil perikanan Jawa Tengah di luar negeri tetap menggeliat di masa pandemi Hal ini ditunjukkan oleh peningka	Positif
8	ntsana bagian paling sedih karna beberapa di antara kita punya janji untuk habis pandemi ketemu ya ternyata sebelum pandemi ini	Netral
9	RadioElshinta Agar tetap fit walaupun sedang DirumahAja olahraga itu penting untuk meningkatkan imun tubuh kita khususnya di masa	Positif
10	yunggichi makan kak ntar sakit jaga kesehatan kak di tengah pandemi ini	Netral
11	VIVAcoid Erick Thohir Kondisi BUMN di Tengah Pandemi Corona 90 Sakit	Negatif
12	AnakLolina Diskusi dengan judul Pemecatan Presiden di tengah Pandemi masih di bilang kebebasan akademik Selesai diskusi lalu muncul	Positif
13	aniesbaswedan Kondisi DKI Jakaa hari ini tak lepas dari kolaborasi berbagai organisasi dan masyarakat dalam menangani pandemi COVID	Negatif
14	ahmadsopandii Buat kamu yang berada didalam atw pun diluar rumah saat pandemi ini harus menjaga kesehatan karn	Positif
15	samimsya ciptohp KuntoDrummer dirgarambe Saat ini beliau yg terbaik menurut saya mungkin berbeda dengan pe	Negatif
16	jokowi Pemerintah saat ini fokus menangani pandemi Covid19 tapi agenda strategis yang berdampak besar bagi kehidupan rakyat harus te	Positif
17	officialJKT48 Target donasi untuk membantu mereka yang terkena PHK di masa pandemi COVID19 telah tercapai Terima kasih atas dukungan	Positif
18	Hilmi28 Presiden sepei Trumph saja ingat akhirat disaat pandemi sepei ini	Netral
19	Rakyat Tak Puas Menteri Pusing Bukti Gagalnya Rezim Atasi Pandemi Muslimah News	Negatif
20	Agar tetap fit walaupun sedang DirumahAja olahraga itu penting untuk meningkatkan imun tubuh kita khususnya di	Positif

id	TEXT	POLARITY
572	pandukusumo NizrinaAyu Kabar gembira Saya menyediakan jasa konsultasi psikolog murah meriah untuk mengatasi gang	Positif

Sumber: olahan peneliti

Dari total 1.250 data dengan kata kunci pandemi dan corona data yang sudah berhasil melewati proses *cleaning* dan sudah diberi label, diketahui bahwa data berlabel positif sebanyak 229 sentimen, data berlabel netral sebanyak 258 sentimen, dan data berlabel negatif sebanyak 763 sentimen.

4.2.1 Dataset

Proses yang pertama dilakukan adalah menggunakan operator *Read Excel* untuk memasukan *dataset*. Adapun data yang dimasukan pada proses *Read Excel* adalah data yang telah melewati tahapan pelabelan. *Dataset* yang berbentuk *Microsoft Excel* pada Rapidminer dibaca dengan menggunakan operator *Read Excel*, seperti tampak pada Gambar 4.4. Pada pembacaan di operator *Microsoft Excel*, ditetapkan bahwa Nomor baris dijadikan *id* dan *Polarity* dijadikan label.



Gambar 4.4 Proses *Read Excel* untuk Memasukan *Dataset*.

Hasil dari pembacaan operator *Read Excel* dihubungkan ke titik *Result* untuk memastikan data yang dibaca oleh operator sesuai dengan *dataset* sebenarnya yang dalam bentuk *Microsoft Excel*. Pada Gambar 4.5 memperlihatkan hasil pembacaan dari operator *Read Excel*, *id* merupakan Nomor baris dan label merupakan *Polarity*.

ExampleSet (1250 examples, 2 special attributes, 1 regular attribute)

Row No.	id	POLARITY	TEXT
1	1	Positif	Pemerintah saat ini fokus menangani pandemi Covid...
2	2	Negatif	Erick Thohir Kondisi BUMN di Tengah Pandemi Coro...
3	3	Netral	Ketika pandemi berdampak kepada penerbit Mas Ag...
4	4	Netral	drkoko28 Tenaga medis sangat diperlukan pada ma...
5	5	Netral	GaekParewa1 Semakin kelihatan kualitas kerjanya R...
6	6	Netral	DRespati2 Pandemi corona hanya melengkapi Dulu ...
7	7	Positif	HumasBKIPM Pasar produk hasil perikanan Jawa Te...
8	8	Netral	ntsana bagian paling sedih karna beberapa di antara...
9	9	Positif	RadioElshinta Agar tetap fit walaupun sedang Dirum...
10	10	Netral	yunggichi makan kak ntar sakit jaga kesehatan kak di...
11	11	Negatif	VIVAcoid Erick Thohir Kondisi BUMN di Tengah Pand...
12	12	Positif	AnakLolina Diskusi dengan judul Pemecatan Preside...
13	13	Negatif	aniesbaswedan Kondisi DKI Jakaa hari ini tak lepas ...
14	14	Positif	ahmadsopandii Buat kamu yang berada didalam atw ...

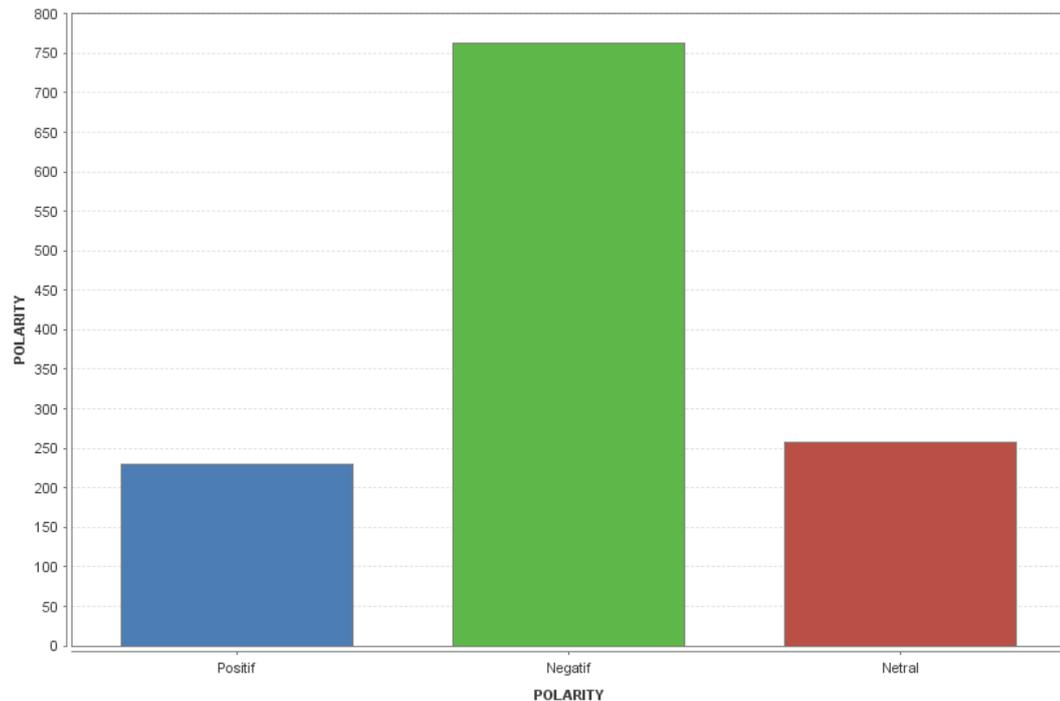
Gambar 4.5 Data Hasil Proses *Read Excel*

Untuk memastikan *dataset* yang digunakan tidak memiliki nilai *missing*, maka dapat dilihat pada statistik dari *dataset* yang digunakan seperti tampak pada Gambar 4.6. Dari nilai statistik yang ditampilkan bahwa *dataset* yang digunakan tidak memiliki nilai *missing*, jumlah total data sebanyak 1.250 *record*.

Column	Type	Count	Sub-categories
No	Integer	1250	625,000
POLARITY	Polynomial	1250	Positif (229), Negatif (763), Netral (258), [1 more]
TEXT	Text	1250	yang sul [] tv 2 (1), @Kemenke [] ong (50), @Kemenke [] wal dong (50), @kaptenA [] []

Gambar 4.6 Statistik Hasil Proses *Read Excel*

Gambar 4.7 menunjukkan grafik batang terhadap *dataset* yang digunakan. Jumlah data 1.250, dengan jumlah data paling sedikit yaitu data positif sebanyak 229 *record*, kemudian data terbanyak adalah data negatif sebanyak 763 *record*. Untuk data netral sebanyak 258 *record*. Pada keseluruhan data yang diinputkan tidak ada *missing value*.



Gambar 4.7 Grafik Hasil Proses *Read Excel*

4.2.2 Process Document

Process document diawali dengan mengubah jenis data pada *dataset* yang dibaca oleh operator *Read Excel*, khususnya pada atribut *Text* yang terdapat angka atau *Nominal*. Diperlukan operator *Nominal to Text* untuk mengubah tipe data dari *Nominal* menjadi *Text* seperti yang ditunjukkan pada Gambar 4.8.



Gambar 4.8 Model Process Document menggunakan RapidMiner.

Tahap pertama dari *Process document* yaitu melakukan *transform cases* atau mengubah huruf pada dataset menjadi huruf kecil. Kemudian dilakukan proses *tokenize* yaitu proses untuk memisahkan kata demi kata pada setiap baris data. Selanjutnya dilakukan proses *filter tokens by length* yaitu proses untuk menyaring kata sesuai dengan

jumlah huruf yang ditentukan. Selanjutnya proses *Filter Stopwords* yaitu untuk menghilangkan kata-kata yang tidak memiliki nilai sentimen. Terakhir dilakukan proses *Stemming* atau pencarian kata dasar atau akar kata. Proses-proses tersebut terdapat pada Gambar 4.9.



Gambar 4.9 *Process Documents*

Kelima proses tersebut dilakukan secara bersamaan pada *proses documents* pada RapidMiner. Proses ini dilakukan karena pada tahapan sebelumnya belum dilakukan proses tersebut.

4.3 Pembobotan

Proses pembobotan dilakukan pada operator *Process Document*. Ketika operator ini digunakan selain di dalamnya ada *preprocessing* juga menghasilkan pembobotan TF *Term Binary*. TF atau *Term Frequency Binary* menunjukkan apakah suatu kata (*term*) terdapat dalam dokumen atau tidak. Jika tidak ada maka beri nilai 0, dan jika ada maka diberi nilai 1.

IDF (*Inverse document Frequency*) menunjukkan hubungan ketersediaan suatu kata (*term*) pada seluruh dokumen. Untuk mendapatkan nilai IDF maka perlu diketahui dulu nilai *df* (total dokumen yang mengandung *term*) pada setiap kata. Maka apabila sudah diketahui nilai *df* selanjutnya dihitung nilai D/df . Dimana nilai *D* merupakan total seluruh dokumen yaitu 1.250. Untuk mendapatkan nilai IDF digunakan rumus $\log\left(\frac{D}{df}\right)$.

Berikut adalah contoh perhitungan IDF dan TF-IDF berdasarkan rumusnya:

Contoh perhitungan IDF dari kata pandemi pada dokumen ke-1 :

$$TF = 1$$

$$D = 1250$$

$$df = 446$$

$$IDF = \log\left(\frac{D}{df}\right)$$

$$= \log\left(\frac{1250}{446}\right)$$

$$= \log(3) = 0,45$$

Contoh perhitungan TF-IDF dari kata pandemi pada dokumen ke-1 :

$$TF - IDF (w) = TF \times (IDF + 1)$$

$$= 1 \times (0,45 + 1)$$

$$= 1,45$$

Maka dari hasil menghitung menggunakan rumus-rumus tersebut dihasilkan nilai seperti yang tertera pada tabel 4.8.

Tabel 4.8 Hasil Proses TF dan IDF

id	aamiin	abad	abai	corona	pandemi	zoom	Jumlah Token
1	0	0	0	0	1	0	9
2	0	0	0	1	1	0	7
3	0	0	0	0	1	0	8
4	0	0	0	0	1	0	8
5	0	0	0	1	0	0	11
6	0	0	0	1	1	0	13
7	0	0	0	0	1	0	10
8	0	0	0	0	2	0	8
9	0	0	0	0	0	0	6
10	0	0	0	0	1	0	9
1250	0	0	0	0	1	0	6
df	11	1	1	576	446	1	
D/df	114	1250	1250	2	3	1250	
IDF	3,06	4,10	4,10	1,34	0,45	4,10	
IDF+1	4,04	5,10	5,10	2,34	1,45	5,10	

Sumber: olahan peneliti

TF-IDF merupakan penggabungan dari kedua metode tersebut dengan cara mengalikan nilai TF dengan nilai DF+1. Dari perhitungan tersebut telah didapatkan hasil seperti pada tabel 4.9.

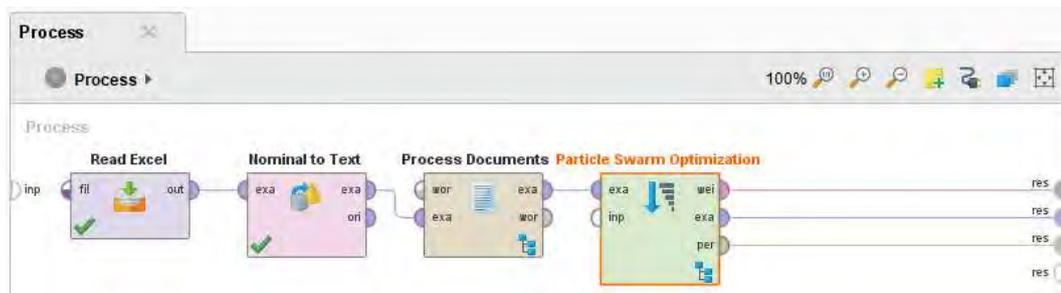
Tabel 4.9 Hasil Proses TF-IDF

id	aamiin	abad	abai		corona		pandemi		zoom
1	0	0	0		0		2,45		0
2	0	0	0		2,34		2,45		0
3	0	0	0		0		2,45		0
4	0	0	0		0		2,45		0
5	0	0	0		2,34		0,00		0
6	0	0	0		2,34		2,45		0
7	0	0	0		0		2,45		0
8	0	0	0		0		4,90		0
9	0	0	0		0		0,00		0
10	0	0	0		0		2,45		0
1250	0	0	0		0		2,45		0

Sumber: olahan peneliti

4.4 Klasifikasi dan Optimasi

Setelah ekstraksi fitur, maka selanjutnya adalah mengklasifikasikan sentimen menggunakan *Support Vector Machine* dengan optimasi *Particle Swarm Optimization* dengan bantuan RapidMiner. Adapun model prosesnya dapat dilihat pada gambar 4.10.



Gambar 4.10 Model Proses

Seperti tampak pada Gambar 4.10, *Particle Swarm Optimization* menggunakan operator *Optimize Weights (PSO)* yang memiliki sub proses didalamnya. Parameter yang digunakan pada *Optimize Weights (PSO)* menggunakan nilai-nilai *default* yang ada pada operator *Optimize Weights (PSO)* atau sesuai pada Tabel 4.10.

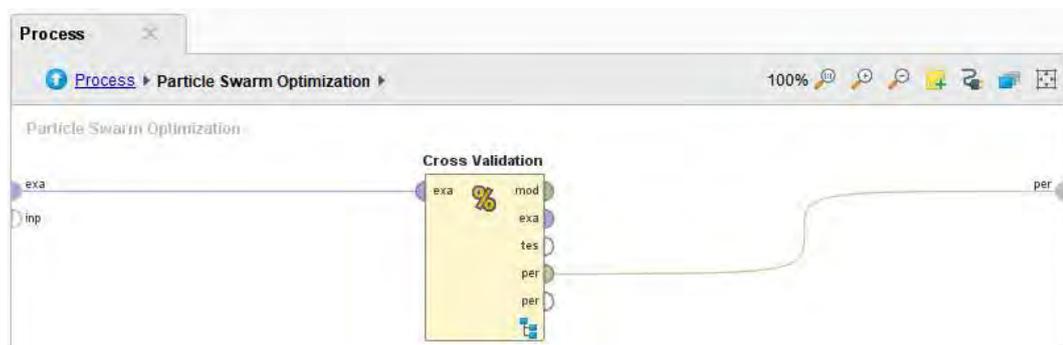
Tabel 4.10 Nilai pada parameter operator *Optimize Weights (PSO)*

No.	Parameter	Value
1.	<i>Population size</i>	5

No.	Parameter	Value
2.	<i>Maximum number of generations</i>	30
3.	<i>Inertia weight</i>	1.0
4.	<i>Local best weight</i>	1.0
5.	<i>Global best weight</i>	1.0
6.	<i>Min weight</i>	0.00
7.	<i>Max weight</i>	1.0

Sumber: olahan peneliti

Operator *Cross Validation* digunakan di dalam operator *Optimize Weights (PSO)* yang digunakan untuk membagi *dataset* untuk menjadi data *training* dan data *testing*. Secara *default*, operator *Cross Validation* akan membagi *dataset* menjadi 80% untuk *data training* dan 20% untuk *data testing*. Gambar 4.11 menunjukkan model proses terhadap operator *Cross Validation* pada operator *Optimize Weights (PSO)*.



Gambar 4.11 Operator *Cross Validation* Pada Operator *Optimize Weights (PSO)*

Parameter yang digunakan pada *Cross Validation* yaitu *number of fold 10* dan *sampling type stratified sampling* atau sesuai dengan yang ada pada Tabel 4.11.

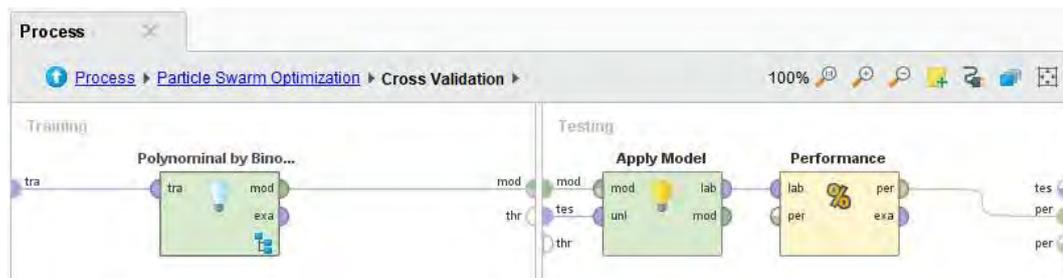
Tabel 4.11 Nilai Parameter Pada Operator *Cross Validation*

No.	Parameter	Value
1.	<i>number of fold</i>	10
2.	<i>sampling type</i>	<i>stratified sampling</i>

Sumber: olahan peneliti

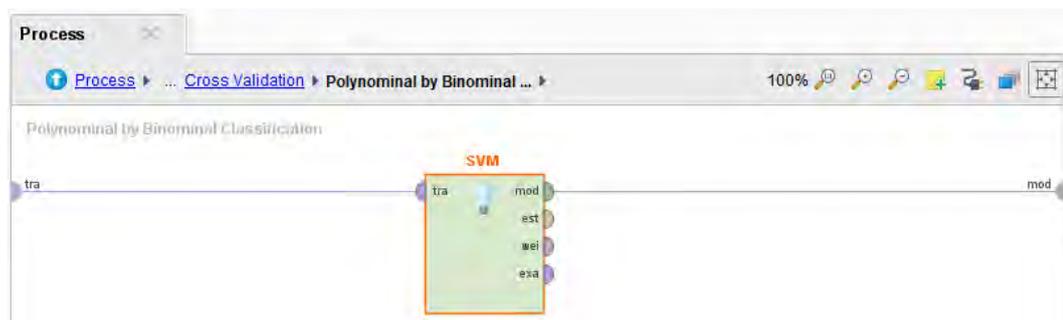
Di dalam operator *Cross Validation*, di tempat sebuah operator *Polynomial by Binominal* untuk melakukan *training* dan dua buah operator yaitu *Apply Model* dan *Performance* untuk melakukan *testing*. Operator *Polynomial by Binominal* ini membuat

model klasifikasi *polinomial* menggunakan klasifikasi *binomial* yang disediakan dalam subprosesnya, parameter *classification strategies* yang digunakan adalah *one against all*. *Apply Model* digunakan untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mentransformasi data. Operator *Performance* digunakan untuk mengevaluasi kinerja hanya untuk tugas klasifikasi, parameter yang digunakan yaitu *main criterion first* dan *accuracy*.



Gambar 4.12 Operator yang Digunakan Pada Operator *Cross Validation*

Pada Gambar 4.13 operator SVM ditempatkan pada operator *Polynomial by Binomial*. Pada penelitian ini parameter yang digunakan *kernel type radial*, dengan kombinasi parameter C, nilai *kernel cache* 200, nilai *max iteration* sebesar 100.000 kemudian nilai *convergence epsilon* yang diubah-ubah untuk mendapatkan akurasi yang maksimal atau seperti yang tampak pada Tabel 4.12.



Gambar 4.13 Operator SVM Pada Operator *Polynomial to Binomial*

Tabel 4.12 Nilai pada parameter operator SVM

No.	Parameter	Value
1.	Kernel type	Radial
2.	Kernel gamma	1.0

No.	Parameter	Value
3.	Kernel cache	200
4.	C	0,0 – 0,5
5.	Convergence epsilon	0,001 - 0,005
6.	Max iterations	100.000
7.	L pos	1,0
8.	L neg	1,0
9.	Epsilon	0,0
10.	Epsilon plus	0,0
11.	Epsilon minus	0,0

Sumber: olahan peneliti

4.5 Validasi dan Evaluasi

Terhadap dataset yang dipergunakan selanjutnya diadakan pengujian model algoritme SVM berbasis PSO dengan menggunakan *Software RapidMiner 9.5* melalui proses *Cross Validation* maka didapatkan hasil klasifikasi analisis dari *tweet* mengenai Pandemi Covid-19 dengan algoritme *Support Vector Machine* berbasis *Particle Swarm Optimization* sebagai berikut.

Berdasarkan beberapa kali percobaan yang dilakukan, diperoleh hasil klasifikasi *sentiment analysis* dari *tweet* mengenai Pandemi Covid-19 dengan algoritme *Support Vector Machine* berbasis *Particle Swarm Optimization* yang terbaik menghasilkan akurasi dengan nilai 93,20% seperti terlihat pada Tabel 4.13.

Tabel 4.13 Nilai Akurasi

No.	C	Convergence Epsilon	Akurasi
1.	0,0	0,001	93,20%
2.	0,0	0,002	93,20%
3.	0,0	0,003	93,20%
4.	0,0	0,004	93,20%
5.	0,0	0,005	93,20%
6.	0,1	0,005	92,99%
7.	0,2	0,005	92,72%
8.	0,3	0,005	93,04%
9.	0,4	0,005	93,04%

No.	C	Convergence Epsilon	Akurasi
10.	0,5	0,005	93,04%
11.	1	0,005	93,20%
12.	2	0,005	93,20%
13.	3	0,005	93,20%
14.	4	0,005	93,20%
15.	5	0,005	93,20%
16.	6	0,005	93,20%
17.	7	0,005	93,20%
18.	8	0,005	93,20%
19.	9	0,005	93,20%
20.	10	0,005	93,20%

Sumber: olahan peneliti

Dari hasil perbandingan pada nilai akurasi seperti pada Tabel 4.11, akurasi tertinggi terdapat pada nilai parameter C yaitu 0,0 dan nilai *Convergence Epsilon* mulai dari 0,001 sampai dengan 0,005 memiliki nilai yang sama terhadap nilai akurasi atau dengan kata lain bahwa *Convergence Epsilon* tidak memiliki pengaruh terhadap nilai akurasi. Tabel 4.12 memperlihatkan *confusion matrix* pada akurasi sebesar 93,20%, dapat dipaparkan sebagai berikut.

1. Data yang diprediksi negatif : ternyata benar-benar negatif sebanyak 756 *record*, dan netral sebanyak 40 *record* serta ternyata positif sebanyak 14 *record*.
2. Data yang diprediksi netral : ternyata benar-benar netral sebanyak 218 *record*, dan ternyata positif sebanyak 24 *record*.
3. Data yang diprediksi positif : ternyata benar-benar positif sebanyak 191 *record*, dan ternyata negatif sebanyak 7 *record*.

Tabel 4.14 *Confusion matrix* pada akurasi tertinggi

	True Negatif	True Netral	True Positif	Class Precision
Pred. Negatif	756	40	14	93,33%
Pred Netral	0	218	24	90,08%
Pred. Positif	7	0	191	96,46%
Class Recall	99,08%	84,50%	83,41%	

Sumber: olahan peneliti

Hasil akurasi menunjukkan rasio prediksi yang benar dengan keseluruhan data sebesar 93,20%. Hasil *class recall true* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan data yang benar negatif sebesar 99,08%. Hasil *class recall true* netral menunjukkan rasio prediksi benar netral dibandingkan dengan keseluruhan data yang benar netral sebesar 84,50%. Hasil *class recall true* positif menunjukkan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif sebesar 83,41%.

Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 93,33%. Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 90,08%. Hasil *class precision Pred.* negatif menunjukkan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif sebesar 96,46%.

Dari penelitian tersebut maka diketahui bahwa analisis sentimen mengenai pandemi covid-19 belum akurat karena opini masyarakat belum dapat diekstraksi polanya. Pemasalahan mengenai pandemi covid-19 dapat dikatakan sesuatu yang baru di masyarakat sehingga cara penyampaian komentar dari masyarakat beragam dan sentimen masyarakat mengenai pandemi covid-19 belum bisa diprediksi. Bagi peneliti selanjutnya disarankan agar tidak mengambil topik-topik insidental untuk analisis sentimen, kecuali menggunakan metode *deep learning*.

BAB IV

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil eksperimen dapat diambil kesimpulan beberapa poin. Diantaranya adalah sebagai berikut:

1. Hasil analisis sentimen pengguna *Twitter* terhadap Pandemi Covid-19 sebanyak 1250 *record* menggunakan Kamus Sastrawi diperoleh yaitu sentimen positif sebanyak 229 *record*, kemudian sentimen negatif sebanyak 763 *record* dan sentimen netral sebanyak 258 *record*.
2. Klasifikasi terhadap analisis sentimen pengguna *Twitter* mengenai Pandemi Covid-19 menggunakan metode *Support Vector Machine* dengan optimasi *Particle Swarm Optimization* diperoleh nilai akurasi sebesar 93,20%. Sentimen yang diprediksi negatif : ternyata benar-benar negatif sebanyak 756 *record*. Sentimen yang diprediksi netral : ternyata benar-benar netral sebanyak 218 *record*. Dan sentimen yang diprediksi positif : ternyata benar-benar positif sebanyak 191 *record*.

5.2 Saran

Adapun saran untuk penelitian berikutnya yang dapat dikembangkan berdasarkan penelitian ini diantaranya yaitu:

1. Menggunakan lebih banyak data sehingga cakupan kosakata lebih luas dan klasifikasi sentimen menjadi lebih akurat.
2. Melakukan analisis sentimen dengan menggunakan metode *deep learning* misalnya dengan dengan algoritme *Deep Convolutional Neural Network*.

DAFTAR PUSTAKA

- Aida, N. R. (2020). Update Virus Corona di Dunia 10 Mei : 4 , 1 Juta Orang Terinfeksi , China Siap Bantu Korut Atasi Pandemi. *Kompas.Com*, 1–8. <https://www.kompas.com/tren/read/2020/05/10/083400665/update-virus-corona-di-dunia-10-mei--4-1-juta-orang-terinfeksi-china-siap?page=all#page3> diakses tanggal 15 Mei 2020 pukul 14.30 wib.
- Cahyono, Y. (2017). Analisis Sentiment pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency. *Jurnal Informatika Universitas Pamulang*, 2(1), 14–19. <https://doi.org/10.32493/informatika.v2i1.1500>
- Darma, I. M. B. S., Perdana, R. S., & Indriati. (2018). Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(3), 998–1007. <http://j-ptiik.ub.ac.id>
- Eka Ratnawati, D., . M., & Muflikhah, L. (2014). Pengembangan Metode Klasifikasi Berdasarkan K-Means Dan LVQ. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 1(1), 1. <https://doi.org/10.25126/jtiik.20141197>
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Gusriani, S., Wardhani, K. D. K., & Zul, M. I. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka). *4th Applied Business and Engineering Conference*, 1(1), 1–7.
- Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, 2016(Sentika), 57–64. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>
- Haryanto, A. T. (2019). Riset Januari 2019: WhatsApp Digilai Netizen Indonesia. *Detikinet*, 1. <https://inet.detik.com/cyberlife/d-4461246/riset-januari-2019-whatsapp-digilai-netizen-indonesia>
- Hasanah, U., Mayangsari, L. R., Pratama, A., & Cholissodin, I. (2016). Perbandingan Metode SVM, FUZZY-KNN, Dan BDT-SVM Untuk Klasifikasi Detak Jantung Hasil Elektrokardiografi. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 3(3), 201–207. <https://doi.org/10.25126/jtiik.201633196>
- Ilmawan, L. B., & Winarko, E. (2015). Aplikasi Mobile untuk Analisis Sentimen pada Google Play. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 9(1), 53–64. <https://doi.org/10.22146/ijccs.6640>
- Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi*, 4(2), 20–27.
- Kristiyanti, D. A. (2015). Analisis Sentimen Review Produk Kosmetik Menggunakan Algoritma Support Vector Machine Dan Particle Swarm Optimization Sebagai Metode Seleksi Fitur. *Seminar Nasional Inovasi & Tren (SNIT) 2015*, 134–141. http://lppm.bsi.ac.id/SNIT2015/BidangA/A22-134-141_2015-SNIT-Dinar_Ajeng_Kristiyanti_ALGORITMA_SUPPORT_VECTOR.pdf
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. 2(11), 4704–4713.
- Meilina, P. (2015). Penerapan Data Mining dengan Metode Klasifikasi Menggunakan

- Decision Tree dan Regresi. *Jurnal Teknologi Universitas Muhammadiyah Jakarta*, 7(1), 11–20. jurnal.ftumj.ac.id/index.php/jurtek
- Merluarini, B., Safitri, D., & Hoyyi, A. (2013). PERBANDINGAN ANALISIS KLASIFIKASIMENGGUNAKAN METODE K-NEAREST NEIGHBOR(K-NN)DAN MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) PADA DATA AKREDITASI SEKOLAH DASAR NEGERIDI KOTA SEMARANG. *Jurnal GAUSSIAN*, 3(2), 313–322. <https://doi.org/10.1007/s11139-020-00300-y>
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 180–184. <https://doi.org/10.25126/jtiik.201743251>
- Mujilawati, S. (2016). PRE-PROCESSING TEXT MINING PADA DATA TWITTER. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 49–56.
- Nurhadi, Z. F. (2017). Model Komunikasi Sosial Remaja Melalui Media Twitter. *Jurnal ASPIKOM*, 3(3), 539–549. <https://doi.org/10.24329/aspikom.v3i3.154>
- Organization, W. H. (2020). Q&A on coronaviruses (COVID-19). *New Scientist*, 246(3280), 56. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses#:~:text=symptoms>
- Pitria, P. (2019). Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naive Bayes. *Undergraduate Theses from JBPTUNIKOMPP*.
- Praptiwi, D. Y. (2018). *Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy*. Universitas Islam Indonesia.
- Pratama, J. A., Suprijadi, Y., & Zulhanif. (2017). The Analisis Sentimen Sosial Media Twitter Dengan Algoritma Machine Learning Menggunakan Software R. *Jurnal Fourier*, 6(2), 85–89. <https://doi.org/10.14421/fourier.2017.62.85-89>
- Pudjajana, A. M., & Manongga, D. (2018). Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan Naive Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 313–318. <https://doi.org/10.24176/simet.v9i1.1922>
- Rachmat, A., & Lukito, Y. (2016). SENTIPOL: Dataset Sentimen Komentar Pada Kampanye PEMILU Presiden Indonesia 2014 dari Facebook Page. *Konferensi Nasional Teknologi Informasi Dan Komunikasi 2017, December*, 218–228.
- RapidMiner. (2012). *RapidMiner Studio Manual*. <https://doi.org/http://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>
- Rohman, A., & Rochcham, M. (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 5(1), 23–29. <https://doi.org/10.37760/neoteknika.v5i1.1379>
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207–217.
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine. *Jurnal Transformatika*, 14(2), 72. <https://doi.org/10.26623/transformatika.v14i2.439>
- Yasmin, P. (2020). Asal Usul Virus Corona Berasal , dari Mana Sebenarnya ? *DetikNews*, 1–5. <https://news.detik.com/berita/d-4966701/asal-usul-virus-corona-berasal-dari-mana-sebenarnya>
- Yislam, & Budi, I. (2016). *Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter*. October.

Yusuf, A., & Priambadha, T. (2013). Support Vector Machines Yang Didukung K-Means Clustering Dalam Klasifikasi Dokumen. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 11(1), 13–16. <https://doi.org/10.12962/j24068535.v11i1.a15>