

**IMPLEMENTASI METODE SUPPORT VECTOR MACHINE DALAM
ANALISIS SENTIMEN PADA DATA ULASAN TWITTER VAKSIN
COVID-19**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Strata-1 Pada Jurusan Teknik Industri Fakultas
Teknologi Industri**



Disusun Oleh :

Nama : Andika Julianto Putra

No. Mahasiswa : 17522246

**PROGRAM STUDI TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2021



FAKULTAS
TEKNIK INDUSTRI

Gedung KH. Mas Mansur
Jl. Kalbarang Km 14,5 Yogyakarta
Telp. (0274) 895287, 898444 ext.2511;
Fax. (0274) 895987

SURAT KETERANGAN PENELITIAN

Nomor : 219/A/Ka.Lab DATMIN/FTI-UII/V/2021

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Andika Julianto Putra
No. Mhs : 17522246
Dosen Pembimbing : Harwati, S.T., M.T.

Telah selesai melaksanakan penelitian yang berjudul "IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* DALAM ANALISIS SENTIMEN PADA DATA ULASAN TWITTER VAKSIN COVID-19" di Laboratorium Data Mining , Prodi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 1 Januari sampai dengan tanggal 28 Februari 2021.

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Dikeluarkan : di Yogyakarta

Tanggal : 31 Mei 2021

Mengetahui,
Kepala Lab. Data Mining

Annisa Uswatun Khasanah, S.T., M.Sc



PERNYATAAN KEASLIAN

Demi Allah saya akui bahwa karya ini adalah karya sendiri kecuali kutipan dan ringkasan yang setiap salah satunya telah saya jelaskan sumbernya. Jika ditemukan dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak kekayaan intelektual maka saya bersedia ijazah yang saya terima untuk ditarik oleh Universitas Islam Indonesia.

Yogyakarta, 28 Mei 2021



Andika Julianto Putra



LEMBAR PENGESAHAN PEMBIMBING

IMPLEMENTASI METODE SUPPORT VECTOR MACHINE DALAM
ANALISIS SENTIMEN PADA DATA ULASAN TWITTER VAKSIN
COVID-19

TUGAS AKHIR



Yogyakarta, 28 Mei 2021

Dosen Pembimbing

Harwati, S.T., M.T.

الجامعة الإسلامية

LEMBAR PENGESAHAN PENGUJI

LEMBAR PENGESAHAN PENGUJI

IMPLEMENTASI METODE SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN PADA DATA ULASAN TWITTER VAKSIN

COVID-19

TUGAS AKHIR

Disusun oleh:

Nama : Andika Julianto Putra

NIM : 17522246

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-I Teknik Industri

Yogyakarta, 12 Juli 2021

Tim Penguji

Harwati, S.T., M.T.

Ketua

Annisa Uswatun Khasanah, S.T., M.B.A., M.Sc.

Anggota 1

Joko Sulistio, S.T., M.Sc.

Anggota 2

Mengetahui,

Ketua Program Studi Teknik Industri

Fakultas Teknologi Industri

Universitas Islam Indonesia



Dr. Fauziah Imawati, S.T., M.M.

HALAMAN PERSEMBAHAN

Dengan nama Allah yang Maha Pengasih dan Maha Penyayang seiring rasa syukur kehadiran Allah SWT, skripsi ini saya persembahkan kepada:

kedua orang tua saya, Ayah Sumaryanto dan Ibu Prima Ernuni Ika Yulia Dewi. Terima Kasih telah mendoakan, mendidik, membina, dan selalu memberikan dukungan dari kecil hingga saat ini. Beliau adalah sosok utama dan segalanya dalam hidup saya yang tidak pernah akan terganti. Betapa segala pengorbanan tidak akan pernah bisa membalas segala bulir keringat dan air mata.

Kakak saya dan sanak keluarga, terima kasih untuk semangat, dukungan, dan motivasi yang selalu diberikan hingga pada tahap ini.

Serta kerabat, sahabat, dan teman-teman saya yang selalu membantu dan hadir menemani hari-hari saya selama masa perkuliahan.

MOTTO



أَسْتَطَعْتُ وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ

“..... Tidak ada kesuksesan melainkan dengan pertolongan Allah SWT. Hanya kepada Allah aku bertawakkal dan hanya kepada-Nya lah aku kembali.”

(QS. Hud: 88)



KATA PENGANTAR



Assalamualaikum Warrah matullahi Wabarakatuh,

**Asyhadu Alla Ilahailallah Wa Asyhadu Anna Muhammadarrasululla h
Allahuma Shalli'ala Muhammad Wa'ala Alihi Washobihi Wasalim,
Alhamdulillahhirra bbil'aa la miin.**

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas berkat rahmat dan nikmat-Nya. Sholawat serta salam senantiasa penulis ucapkan kepada Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya yang telah berjuang dan membimbing kita keluar dari zaman jahilliyah menuju zaman dengan penuh ilmu pengetahuan.

Atas rahmat dan nikmat Allah SWT, tugas akhir yang berjudul **“IMPLEMENTASI METODE SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN PADA DATA ULASAN TWITTER VAKSIN COVID-19”** dapat diselesaikan dengan baik. Dalam penelitian tugas akhir ini penulis mendapat pengetahuan, bimbingan, dukungan, dan sara dari berbagai pihak, oleh karena itu, penulis ingin mengucapkan terimakasih kepada:

1. Bapak Hari Purnomo, Prof., Dr., Ir., M.T. selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Dr. Taufiq Immawan S.T., M.M. selaku Ketua Jurusan Strata-1 Teknik Industri Universitas Islam Indonesia.
3. Ibu Harwati S.T., M.T. selaku Dosen Pembimbing tugas akhir Teknik Industri Universitas Islam Indonesia.
4. Kedua orang tua atas segala doa, dukungan baik moril maupun materil dan semangat yang diberikan.
5. Kedua kakak saya yang telah membantu memberikan solusi atas masalah yang ada.
6. Sahabat semasa kuliah terimakasih atas dukungan yang tiada hentinya diberikan.
7. Teman-teman serta sahabat Teknik Industri Fakultas Teknologi Industri

Universitas Islam Indonesia.

8. Serta teman-teman SMA yang telah memberi dukungan dalam pengerjaan skripsi ini.

Namun tidak lepas dari semua itu, penulis menyadari sepenuhnya bahwa ada kekurangan baik dari bahasa, isi dan format laporan yang telah disusun penulis. Oleh karena itu kritik dan saran untuk memperbaiki laporan penelitian ini. Semoga laporan ini dapat bermanfaat bagi semua pihak. Aaamiin.

Wassalamu'alaikum Warrahmatullahi Wabarakatuh.

Yogyakarta, 28 Mei 2021



Andika Julianto Putra



ABSTRAK

Penyakit corona virus atau *Corona Virus Disease-19* (COVID-19) adalah infeksi saluran pernapasan yang disebabkan oleh jenis virus corona. Nama lain dari penyakit ini adalah *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-COV2). Kasus Covid-19 pertama kali dilaporkan di kota Wuhan, Provinsi Hubei, Tiongkok, pada Desember 2019. Dalam beberapa bulan saja, penyebaran penyakit ini telah menyebar ke berbagai negara, baik di Asia, Amerika, Eropa, dan Timur Tengah serta Afrika. Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia atau *World Health Organization* (WHO) mendeklarasikan penyebaran Covid-19 dikategorikan sebagai pandemi. Sebanyak 1,2 juta dosis vaksin Covid-19 buatan Sinovac tiba di Indonesia pada tanggal 6 Desember 2020. Dasar pemerintah dalam memilih vaksin tersebut antara lain memenuhi unsur aman, efektif, dan mutu yang terjamin oleh lembaga berwenang. Pada tanggal 13 Januari 2021 program vaksinasi vaksin Covid-19 mulai dilaksanakan dengan Presiden Joko Widodo sebagai penerima yang pertama. Beberapa masyarakat berupaya untuk menanggulangi dampak psikis akibat pandemi Covid-19 adalah dengan sekedar menyalurkan keluh kesahnya atau pun saling berbagi informasi melalui media sosial salah satunya Twitter. Berdasarkan cuitan pada media sosial Twitter penulis ingin mencari bagaimana pendapat masyarakat terutama di Indonesia tentang vaksin Covid-19. Apa faktor-faktor yang sering dikeluhkan pengguna Twitter dan apa saja yang dirasakan. Ulasan yang digunakan pada penelitian ini berjumlah 3131 yang diklasifikasikan menjadi kelas sentiment positif dan negatif. Pengklasifikasian data *tweet* menggunakan algoritma *support vector machine* (SVM). Didapatkan hasil akurasi terbaik dengan menggunakan perbandingan data *training* 90% dan data *testing* 10% sebesar 94.88% untuk metode SVM dengan kernel linear. Kemudian, metode asosiasi teks pada kelas sentiment positif diantaranya terkait sehat, kesehatan, aman, protokol, dukung, halal, patuh, lawan, program, ekonomi. Sedangkan pada kelas sentimen negatif diantaranya terkait takut, virus, suntik, masyarakat, sinovac, lupa, penyebaran, proses, efek, salah. Hasil uasan negatif tersebut dianalisis sebab dan akibatnya menggunakan diagram pohon untuk pemecahan masalah.

Kata kunci: analisis sentiment, *support vector machine*, *twitter*, vaksin Covid-19, asosiasi kata, diagram pohon.

Daftar Isi

PERNYATAAN KEASLIAN	iii
LEMBAR PENGESAHAN PEMBIMBING.....	iv
HALAMAN PERSEMBAHAN	vi
MOTTO	vii
KATA PENGANTAR	viii
ABSTRAK.....	x
Daftar Isi	xi
Daftar Tabel	xiii
Daftar Gambar.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	7
1.6 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA	9
2.1 Kajian Induktif	9
2.2 Kajian Deduktif.....	20
2.2.1 Imunisasi.....	20
2.2.2 Vaksin	20
2.2.3 Twitter	21
2.2.4 Data Mining	23
2.2.5 Machine Learning	25
2.2.6 <i>Pre-Processing</i>	25
2.2.7 Sentiment Analysis	26
2.2.8 Classification.....	27
2.2.9 Ukuran Evaluasi Model Klasifikasi.....	29
2.2.10 Fitur dan Pembobotan	30
2.2.11 <i>Support Vector Machine</i>	31
2.2.12 Asosiasi Kata.....	32
2.2.13 Diagram Pohon	32

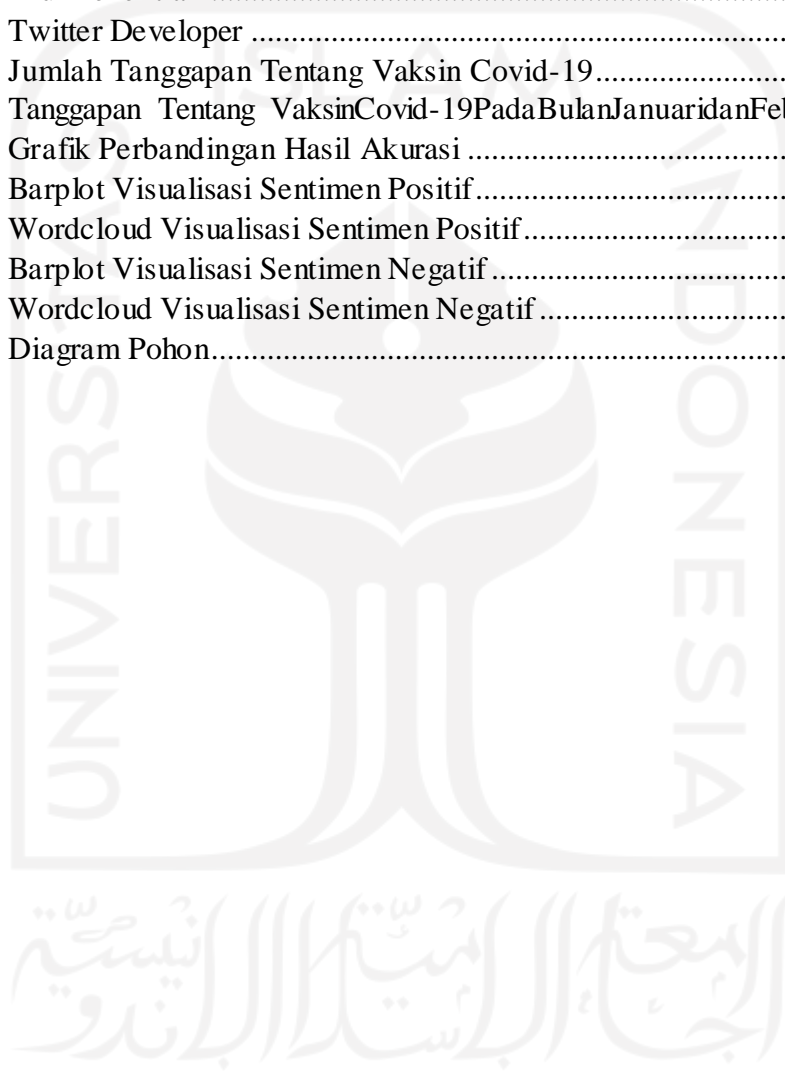
BAB III METODE PENELITIAN	33
3.1 Objek dan Subjek Penelitian	33
3.2 Populasi dan Sampel Penelitian	33
3.3 Pengumpulan Data	33
3.3.1 Metode Pengumpulan Data	33
3.3.2 Jenis Data	34
3.3.3 Variabel dan Definisi Operasional Variabel	34
3.4 Metode Analisis Data	35
3.5 Alur Penelitian.....	35
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA	39
4.1 Pengumpulan Data	39
4.2 Pengolahan Data.....	40
4.2.1 <i>Pre-processing Data</i>	40
4.2.2 Pelabelan dan Pembobotan Kelas Sentimen	43
4.2.3 Pembagian Data <i>Training</i> dan Data <i>Testing</i>	46
4.2.4 Klasifikasi dengan <i>Support Vector Machine</i>	48
4.2.5 Visualisasi dan Asosiasi.....	54
4.2.6 Interpretasi Hasil.....	62
BAB V PEMBAHASAN	64
5.1 Gambaran Umum Persepsi masyarakat.....	64
5.2 Hasil Penerapan Metode <i>Support Vector Machine</i> Dalam Proses Klasifikasi Ulasan Sentimen	65
5.3 Hasil Visualisasi dan Asosiasi Kata	66
5.4 Penjelasan Diagram Pohon.....	67
BAB VI PENUTUP	71
6.1 Kesimpulan.....	71
6.2 Saran.....	72
Daftar Pustaka	73
Lampiran	78
Lampiran 1. Scraping Data	78
Lampiran 2. Preprocessing.....	78
Lampiran 3. Pelabelan.....	81
Lampiran 4. <i>Script R</i> Visualisasi dan Asosiasi Kata	82
Lampiran 5. <i>Output SVM</i>	85

Daftar Tabel

Tabel 2. 1 Analisis Penelitian Terdahulu	11
Tabel 2. 2 Posisi Penelitian.....	16
Tabel 2. 4 Model Klasifikasi.....	29
Tabel 3. 1 Variabel Penelitian.....	34
Tabel 4. 1 Contoh Data Untuk Preprocessing.....	40
Tabel 4. 2 Menghapus URL Pada Data	41
Tabel 4. 3 Menghapus Tanda Baca Pada Data.....	41
Tabel 4. 4 Menyeragamkan Huruf Menjadi Huruf Kecil (Lower-Case)	42
Tabel 4. 5 Tokenizing	42
Tabel 4. 6 Filtering.....	42
Tabel 4. 7 Stemming	43
Tabel 4. 8 Pelabelan Kelas Sentimen Pada Kalimat	44
Tabel 4. 9 Perhitungan Skor Sentimen	45
Tabel 4. 10 Perbandingan Data Training dan Data Testing.....	47
Tabel 4. 11 Perbandingan Data Training dan Data Testing.....	47
Tabel 4. 12 Perbandingan Data Training dan Data Testing.....	47
Tabel 4. 13 Perbandingan Data Training dan Data Testing.....	47
Tabel 4. 14 Perbandingan Data Training dan Data Testing.....	48
Tabel 4. 15 Perbandingan Akurasi Metode Kernel.....	49
Tabel 4. 16 Tahap Melakukan Analisis SVM.....	49
Tabel 4. 17 Hasil 1	51
Tabel 4. 18 Hasil 2	51
Tabel 4. 19 Hasil 3	52
Tabel 4. 20 Hasil 4	53
Tabel 4. 21 Hasil 5	53
Tabel 4. 22 Asosiasi Kata Positif.....	56
Tabel 4. 23 Asosiasi Kata Negatif	60
Tabel 5. 1 Pembahasan Diagram Pohon	67

Daftar Gambar

Gambar 1. 1 Jumlah Pengguna Media Social di Indonesia	3
Gambar 1. 2 Platforms Media Sosial Yang Paling Aktif	4
Gambar 2. 1 Tahapan dalam Knowledge Discpvery From Database (KDD)	24
Gambar 2. 2 Klasifikasi Menggunakan Model Classifier (KDD)	28
Gambar 2. 3 Klasifikasi Menggunakan Model Classifier (KDD)	31
Gambar 3. 1 Alur Penelitian	36
Gambar 4. 1 Twitter Developer	39
Gambar 4. 2 Jumlah Tanggapan Tentang Vaksin Covid-19	44
Gambar 4. 3 Tanggapan Tentang VaksinCovid-19PadaBulanJanuaridanFebruari.	46
Gambar 4. 4 Grafik Perbandingan Hasil Akurasi	54
Gambar 4. 5 Barplot Visualisasi Sentimen Positif	55
Gambar 4. 6 Wordcloud Visualisasi Sentimen Positif	56
Gambar 4. 7 Barplot Visualisasi Sentimen Negatif	59
Gambar 4. 8 Wordcloud Visualisasi Sentimen Negatif	60
Gambar 4. 9 Diagram Pohon.....	63



BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit corona virus atau *Corona Virus Disease-19* (COVID-19) adalah infeksi saluran pernapasan yang disebabkan oleh jenis virus corona. Nama lain dari penyakit ini adalah *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-COV2). Kasus Covid-19 pertama kali dilaporkan di kota Wuhan, Provinsi Hubei, Tiongkok, pada Desember 2019. Dalam beberapa bulan saja, penyebaran penyakit ini telah menyebar ke berbagai negara, baik di Asia, Amerika, Eropa, dan Timur Tengah serta Afrika. Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia atau *World Health Organization* (WHO) mendeklarasikan penyebaran Covid-19 dikategorikan sebagai pandemi. (USK, 2020).

Menjelang akhir tahun 2020, terdapat berbagai macam kandidat vaksin Covid-19 dengan tingkat keamanan dan keefektifan yang tinggi mulai bermunculan. Beberapa negara pun telah mengeluarkan izin penggunaan darurat (EUA) pemakaian vaksin. Namun disisi lain, terdapat beberapa kelompok antivaksin atau masyarakat yang ragu untuk disuntik vaksin. Hal ini dibuktikan dari berbagai literatur hasil penelitian yang telah diterbitkan oleh berbagai macam journal international yang menyebutkan bahwa keraguan dan penolakan vaksin sudah menjadi fenomena yang ditemukan bahkan jauh sebelum pandemic Covid-19 terjadi. (Putri G. S., 2020)

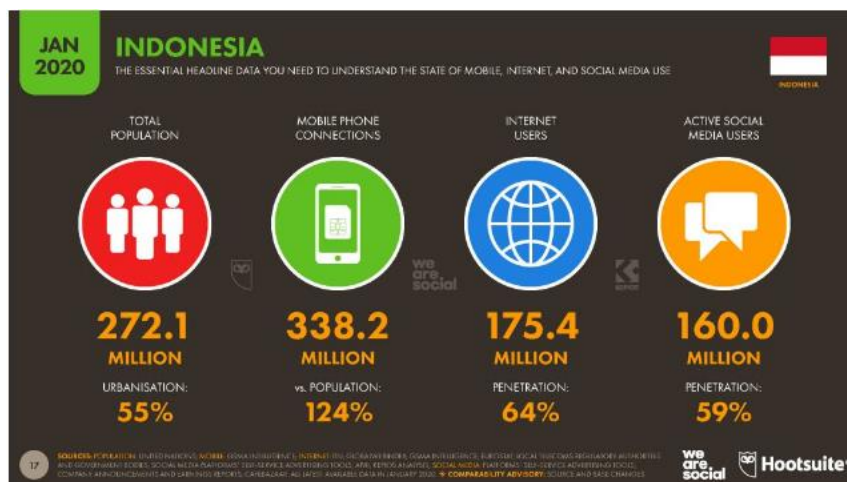
Rencana imunisasi vaksin Covid-19 dianggap terlalu buru-buru mengingat pada saat itu belum ada satupun vaksin yang dinyatakan lolos semua tahapan uji.

Beberapa kolegium dokter bahkan telah melayangkan surat kepada pemerintah untuk meninjau kembali rencana ini (Rahayu, 2020). Bahkan terdapat beberapa tokoh masyarakat yang menolak vaksin seperti Politikus Partai Demokrasi Indonesia Perjuangan (PDIP) Ribka Tjiptaning, dan musisi Gede Arti Astina atau yang disebut Jerinx (Tribunnews, 2021).

Sebanyak 1,2 juta dosis vaksin Covid-19 buatan Sinovac tiba di Indonesia pada tanggal 6 Desember 2020. Dasar pemerintah dalam memilih vaksin tersebut antara lain memenuhi unsur aman, efektif, dan mutu yang terjamin oleh lembaga berwenang. Pada tanggal 13 Januari 2021 program vaksinasi vaksin Covid-19 mulai dilaksanakan dengan Presiden Joko Widodo sebagai penerima yang pertama. (Azizah, 2021)

Informasi vaksinasi serta tata cara pencegahan virus ini telah tersebar di berbagai media sosial. Dengan bantuan teknologi informasi penyampaian informasi menjadi lebih efektif, dan efisien.

Teknologi informasi merupakan bagian dari kehidupan manusia saat ini. Dalam berbagai bidang, teknologi berperan penting untuk menunjang setiap aktivitas kehidupan manusia serta mampu mempermudah setiap keperluan hidup manusia sehari-hari. Salah satu contoh kemajuan teknologi informasi saat ini adalah dengan hadirnya internet. Dengan internet, segala informasi yang diinginkan dapat dengan mudah dan cepat didapatkan (Napitupulu, 2017). Kemajuan teknologi internet telah menjadi simbol cara berkomunikasi secara bebas, tanpa dibatasi oleh ruang, jarak dan waktu. Ditunjang oleh berbagai kelebihan dan keunggulan dari internet, antara lain adalah biaya koneksi internet yang relatif terjangkau dan ketersediaan informasi yang tidak terbatas, internet saat ini menjadi salah satu kebutuhan utama masyarakat dalam memenuhi segala kebutuhan yang terkait dengan informasi. Terbukti pengguna internet di Indonesia mencapai 196,7 juta jiwa terhitung sampai juni 2020 (Pratama, 2020). Salah satu pemanfaatan penggunaan internet adalah media sosial.

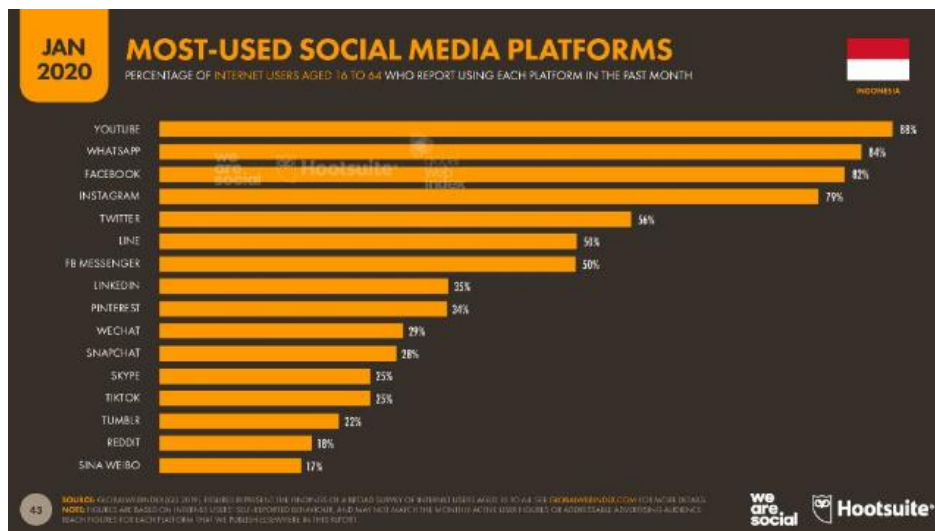


Gambar 1. 1 Jumlah Pengguna Media Social di Indonesia

(Riyanto, 2020)

Pada gambar 1.1 dapat diketahui total pengguna media social di Indonesia mencapai 160 juta jiwa menggunakan atau setara dengan 59% dari jumlah total penduduk Indonesia, dengan pengguna berbasis *mobile* nya mencapai 338,2 juta pengguna. Pengguna media sosial di Indonesia paling banyak berada pada usia produktif yaitu usia 18-34 tahun baik laki-lai maupun perempuan.

Media social seperti *Twitter*, *Facebook*, dan *Youtube* merupakan beberapa media perangkat komunikasi terpopuler di masyarakat saat ini (Aliandu, 2012). Salah satu media jejaring sosial yang telah dimanfaatkan belakangan ini adalah *Twitter* dimana sudah sangat umum di beberapa negara seperti Singapura, Jerman, dan Amerika. Menurut data dari laporan finansial *Twitter* kuartal ke3 tahun 2019, pengguna aktif harian di platform *Twitter* dicatat meningkat 17 persen, ke angka 145 juta pengguna, dan Indonesia menjadi salah satu negara yang pertumbuhan pengguna aktif harian *Twitter* paling besar (Clinton, 2019).



Gambar 1. 2 Platforms Media Sosial Yang Paling Aktif
(Riyanto, 2020)

Twitter merupakan layanan media sosial yang memudahkan penggunaanya untuk berkomunikasi dengan pengguna lainnya dengan mengirimkan pesan yang memiliki kapasitas 280 karakter dengan sebutan *tweet* (Maulana, 2017). Menurut Riyanto per Januari 2020 *Twitter* menduduki peringkat ke lima pada kategori *social network* dengan jumlah sebanyak 56. *Twitter* seringkali digunakan untuk mengungkapkan emosi mengenai sesuatu hal baik, seperti memuji ataupun mencela dalam bentuk emosi. Emosi itu sendiri dapat dikelompokkan menjadi emosi positif dan emosi negatif. Emosi manusia dapat dikategorikan menjadi lima emosi dasar yaitu cinta, senang, sedih, marah, dan takut. Emosi cinta dan senang merupakan emosi positif, sedangkan emosi sedih, marah, dan takut merupakan emosi negatif (Shaver et al., 2001).

Kelebihan *Twitter* dibanding media sosial lainnya yaitu membantu penyebaran informasi secara lebih cepat yang kemudian akan menjadi sebuah topik yang dibahas oleh para penggunanya. Hal ini memudahkan masyarakat mendapatkan informasi secara *real time* dan *up to date* karena berita yang di *update* setiap saat oleh. *Twitter* juga memudahkan menghubungkan dengan pengguna lain tanpa harus mengikuti satu sama lain ketika ingin mengetahui nama akun yang di *mention*.

Menjelang rencana pemerintah dalam distribusi vaksin Covid-19 di seluruh Indonesia, masyarakat dapat memanfaatkan *Twitter* untuk menilai pandangan

tersebut dan dapat diungkapkan dalam bentuk emosi. Emosi *tweet* para pengguna *Twitter* dapat dikenali dengan analisa opini atau sentiment (*opinion analysis* atau *sentiment analysis*). Analisis sentiment juga dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif terhadap suatu tokoh tertentu (Pang & Lee, 2008).

Dari permasalahan diatas, dikaji lebih lanjut mengenai persepsi masyarakat tentang vaksin Covid-19. Persepsi tersebut nantinya dibagi menjadi persepsi positif dan negatif. Evaluasi ulasan tersebut diolah menggunakan analisis sentiment menggunakan pendekatan klasifikasi. Metode klasifikasi *Super Vector Machine* (SVM) dipilih karena kinerja yang dihasilkan bagus, tingkat keakuratan yang dinilai cukup tinggi untuk pengklasifikasian data, dan *error rate* yang dihasilkan minimum (Fathan Hidayatullah & Sn, 2014). Kemudian kelebihan dari metode ini ialah dapat menentukan *hyperplane* atau bidang pemisah dengan memilih bidang dengan optimal margin maka generalisasi pada SVM dapat terjaga dengan sendirinya, tingkat generalisasi pada SVM tidak dipengaruhi oleh jumlah data latih, dengan menentukan parameter *soft margin*, *noise* dapat dikontrol sehingga makin besar parameter *soft margin*, makin besar pula pinalti yang dikenakan pada kesalahan pada klasifikasi sehingga proses pelatihan semakin ketat (Faradhilla et al., 2016).

Selanjutnya dari proses klasifikasi, dilakukan ekstraksi dan eksplorasi informasi pada setiap persepsi positif dan negatif menggunakan asosiasi kata. Pada hasil asosiasi dan visualisasi negatif di analisis menggunakan diagram pohon untuk mengetahui penyebab persepsi negatif masyarakat. Penelitian ini diharapkan dapat menyajikan klasifikasi analisis sentimen yang baik dan tepat sehingga dapat memberikan informasi yang berguna untuk meningkatkan performa dalam distribusi vaksin Covid-19 di Indonesia

1.2 Rumusan Masalah

Meskipun presiden Joko Widodo telah melakukan vaksinasi di depan masyarakat Indonesia, masih ada persepsi negatif masyarakat mengenai vaksin Covid-19. Maka dari itu dapat diambil rumusan masalah yaitu:

1. Bagaimana segmentasi masyarakat Indonesia berdasarkan persepsi terkait dengan vaksin Covid-19?
2. Bagaimana ketepatan klasifikasi *machine learning* dari metode algoritma *Support Vector Machine* pada *Twitter* mengenai vaksin Covid-19?
3. Berdasarkan hasil visualisasi dan asosiasi negatif, pokok masalah apa saja yang dapat terjadi dari hasil tersebut?

1.3 Batasan Masalah

Adapun penelitian ini terdapat beberapa batasan masalah yang diberikan adalah sebagai berikut.

1. Data yang dianalisis hanya berasal dari ulasan *Twitter* terkait vaksin Covid-19.
2. Data diambil dari bulan Januari sampai bulan Februari.
3. Data diambil hanya *tweets* menggunakan bahasa Indonesia.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang ada maka tujuan dilakukan penelitian ini adalah:

1. Menganalisa segmentasi masyarakat Indonesia berdasarkan persepsi terkait dengan vaksin Covid-19.
2. Menghitung ketepatan klasifikasi *machine learning* dari metode algoritma *Support Vector Machine* pada *Twitter* mengenai vaksin Covid-19.

3. Mengidentifikasi pokok masalah yang dapat terjadi berdasarkan hasil visualisasi dan asosiasi negatif.

1.5 Manfaat Penelitian

Manfaat yang diinginkan dalam penelitian ini adalah:

1. Pengklasifikasikan data ulasan tentang vaksin Covid-19 dapat memudahkan pihak pemerintah dalam mengetahui persepsi masyarakat dalam bentuk opini negatif dan opini positif, sehingga dapat dijadikan sebagai acuan dalam upaya distribusi vaksin Covid-19.
2. Menjadi referensi bagi penelitian berikutnya yang relevan mengenai media sosial khususnya *Twitter*.

1.6 Sistematika Penulisan

Untuk menghindari kesalahan dalam penyusunan penulisan tugas akhir, penulis membagi sistematika penulisan kedalam 6 bab, yaitu:

BAB I PENDAHULUAN

Pendahuluan berisi tentang latar belakang dalam penelitian, rumusan masalah, tujuan penelitian, batasan masalah dalam penelitian, manfaat penelitian, dan juga sistematika penulisan.

BAB II KAJIAN LITERATUR

Kajian literatur dibagi menjadi 2, yaitu kajian deduktif serta kajian induktif. Dalam kajian literatur ini terdapat uraian tentang penelitian-penelitian masa lalu yang berkaitan dengan penelitian ini serta teori materi pengetahuan yang mendukung dalam penelitian.

BAB III METODE PENELITIAN

Dalam bab metode penelitian terdapat hal-hal yang akan disampaikan, yaitu objek penelitian, pengumpulan data, cara dalam pengolahan data, alur penelitian.

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

pada bab pengumpulan dan pengolahan data akan dijelaskan tentang bagaimana cara pengumpulan data dan bagaimana pengolahan data sehingga hasilnya dapat digunakan pada bab selanjutnya.

BAB V PEMBAHASAN

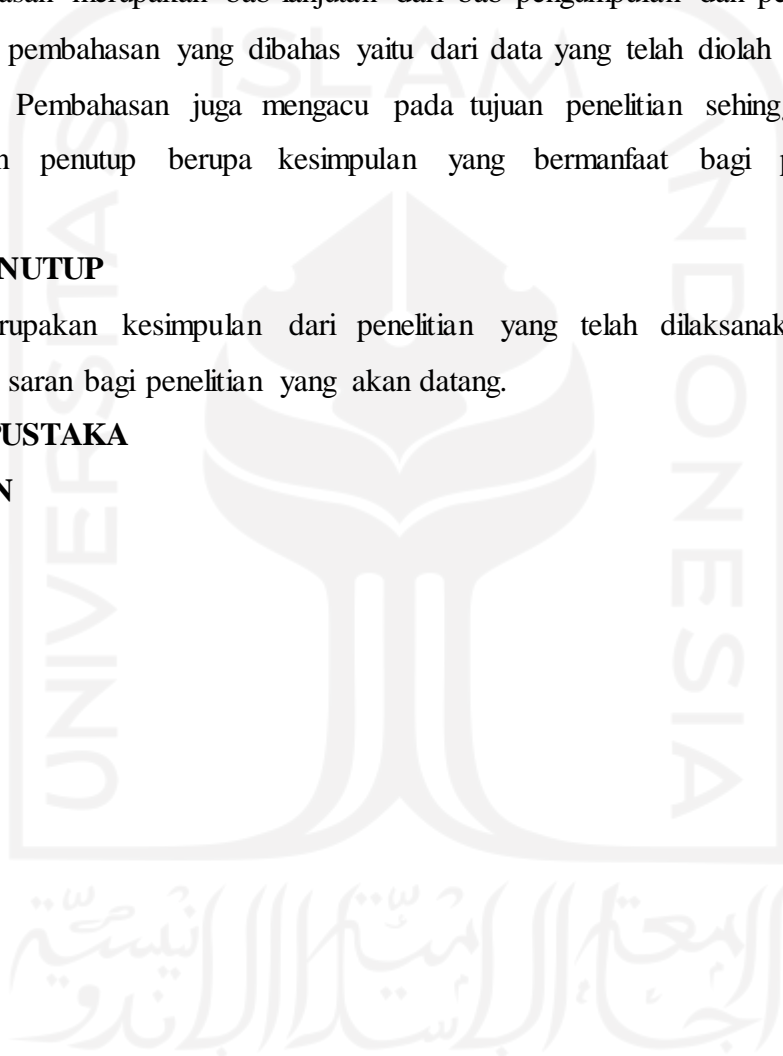
Bab pembahasan merupakan bab lanjutan dari bab pengumpulan dan pengolahan data dimana pembahasan yang dibahas yaitu dari data yang telah diolah pada bab sebelumnya. Pembahasan juga mengacu pada tujuan penelitian sehingga dapat menghasilkan penutup berupa kesimpulan yang bermanfaat bagi penelitian selanjutnya.

BAB VI PENUTUP

Penutup merupakan kesimpulan dari penelitian yang telah dilaksanakan serta memberikan saran bagi penelitian yang akan datang.

DAFTAR PUSTAKA

LAMPIRAN



BAB II

TINJAUAN PUSTAKA

2.1 Kajian Induktif

Penelitian mengenai implementasi analisis sentiment dengan menggunakan metode *Support Vector Machine* bukanlah yang pertama kali dilakukan. *Support Vector Machine* sendiri merupakan salah satu dari sekian banyak metode. Metode-metode lain yang sering digunakan ialah *Maximum Entropy*, dan *Naïve Bayes Classifier*. Pada penelitian ini menggunakan metode *Support Vector Machine* karena metode ini mencari *hyperplane* terbaik untuk memisahkan kedua kelas. Salah satu penerapan metode *Support Vector Machine* dalam bidang kesehatan adalah mengetahui klasifikasi penyakit dengan ciri-ciri yang sama untuk mencegah timbulnya penyakit yang sama, penelitian ini pernah dilakukan oleh Aprilla et al., (2018). Kemudian penerapan lainnya seperti review ulasan pengguna aplikasi, penelitian ini pernah dilakukan oleh Fanani (2017).

Penerapan-penerapan lainnya seperti membantu riset atas opini publik sentimen mengenai objek atau subjek tertentu yang disampaikan di *social media* khususnya *Twitter*, pada topik ini pernah dilakukan oleh Putranti & Winarko (2014), Fathan Hidayatullah & Sn (2014), Novantirani et al. (2015), Buntoro (2016), Faradhillah et al., (2016), Pintoko & Lhaksmana (2018), Fanani (2017), Rofiqoh et al., (2017).

Berdasarkan penelitian menggunakan metode *Support Vector Machine*, dikatakan bahwa akurasi yang didapat dari metode tersebut cukup meyakinkan. Hal ini dibuktikan dari penelitian yang melakukan perbandingan metode seperti yang dilakukan oleh Putranti & Winarko (2014), Fathan Hidayatullah & Sn (2014),

Buntoro (2016), Prakoso et al. (2017), Fanani (2017) Berikut merupakan pemetaan penelitian terdahulu mengenai penelitian analisis sentiment:



Tabel 2. 1 Analisis Penelitian Terdahulu

Judul, Penulis, Tahun	Metode	Hasil Penelitian
Analisis Sentimen <i>Twitter</i> untuk teks Berbahasa Indonesia dengan <i>Maximum Entropy</i> dan <i>Support Vector Machine</i> (Putranti & Winarko, 2014)	<i>Maximum Entropy</i> dan <i>Support Vector Machine</i>	Hasil implementasi menunjukkan ME digunakan untuk POS <i>tagger</i> menghasilkan akurasi 81.67% dan SVM menghasilkan akurasi 86.81% pada pengujian <i>7-fold cross validation</i> untuk tipe kernel <i>sigmoid</i> .
Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada <i>Twitter</i> (Fathan Hidayatullah & Sn, 2014)	<i>Naïve Bayes Classifier</i> (NBC) dan <i>Support Vector Machine</i>	Metode <i>Support Vector Machine</i> (SVM) menghasilkan akurasi performansi yang lebih baik daripada metode <i>Naïve Bayes Classifier</i> (NBC).

Judul, Penulis, Tahun	Metode	Hasil Penelitian
Analisis sentiment pada Twitter Untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota (Novantirani et al, 2015)	<i>Support Vector Machine</i>	Analisis sentiment terhadap data Twitter mengenai penggunaan transportasi umum darat dalam kota dapat dilakukan dengan metode <i>Support Vector Machine</i> , dengan akurasi mencapai 78.12% pada dataset transjakarta.
Analisis Sentimen <i>Hatespeech</i> Pada Twitter (Buntoro, 2016)	<i>Support Vector Machine</i> dan <i>Naïve Bayes Classifier</i>	Hasil akurasi tertinggi didapatkan saat menggunakan metode klasifikasi Support Vector Machine dengan nilai rata-rata akurasi mencapai 66.6%.

Judul, Penulis, Tahun	Metode	Hasil Penelitian
Eksperimen Sistem Klasifikasi Sentiment Twitter Pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin (Faradhillah et al., 2016)	<i>Support Vector Machine, Naïve Bayes Classifier</i>	Hasil akurasi tertinggi didapatkan saat menggunakan metode <i>Support Vector Machine</i> sebesar 79.81% .
Analisis Sentiment Pada Twitter Terhadap Jasa Transportasi Online Di Indonesia (Pintoko & Lhaksmana, 2018)	<i>Naïve Bayes Classifier</i>	Proses klasifikasi menggunakan metode <i>Naïve Bayes Classifier</i> menunjukkan hasil akurasi sebesar 86.80%
Analisis Menggunakan Sentimen <i>Support Vector Machine</i> dan <i>Maximum Entropy</i> (Prakoso et al., 2017)	<i>Support Vector Machine</i> dan <i>Maximum Entropy</i>	Nilai akurasi yang diperoleh dari metode Maxent adalah rata-rata sebesar 74%, sedangkan

Judul, Penulis, Tahun	Metode	Hasil Penelitian
		metode SVM diperoleh rata-rata sebesar 75%.
Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler di Indonesia Pada Twitter (Rofiqoh et al., 2017)	<i>Support Vector Machine</i> (SVM)	Nilai akurasi yang diperoleh dari analisis sentimen sebesar 79%
Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan (Fanani, 2017)	<i>Support Vector Machine, Naïve Bayes Classifier</i>	Penelitian ini menunjukkan performa sistem klasifikasi review SPAM dan HAM dengan algoritma SVM menghasilkan akurasi lebih baik dibandingkan dengan menggunakan algoritma NBC

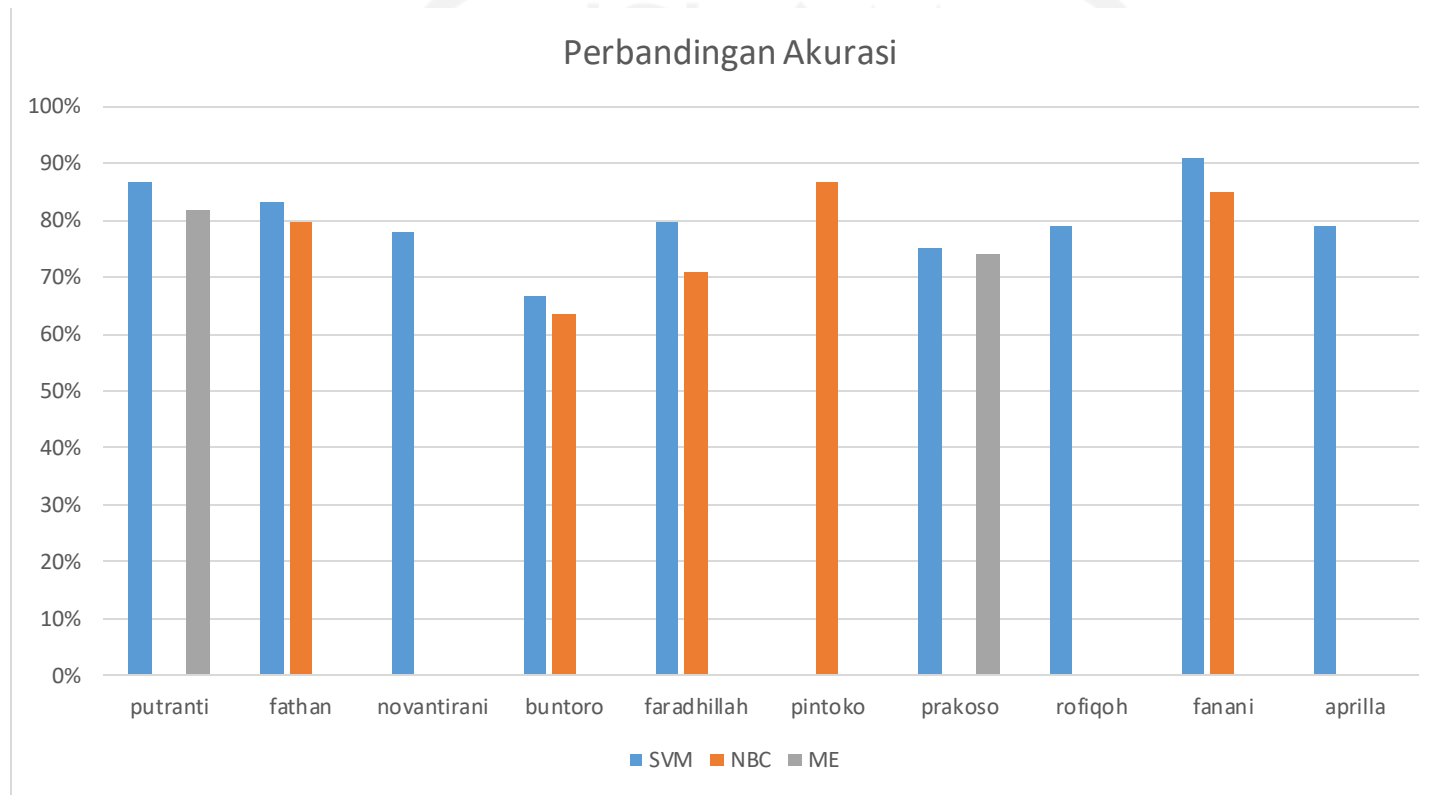
Judul, Penulis, Tahun	Metode	Hasil Penelitian
Klasifikasi penyakit skizofrenia dan episode depresi pada pasien gangguan kejiwaan dengan menggunakan metode <i>Support Vector Machine</i> (Aprilla et al., 2018)	<i>Support Vector Machine</i>	Data yang dimiliki berupa data rekam medis pasien dengan penyakit skizofrenia hebefrenik dan depresi berat dengan masing-masing sebanyak 200 data dengan ratio perbandingan data train dan data test 80%:20% menghasilkan tingkat akurasi sebesar 79%.

Berikut merupakan posisi penelitian yang dilakukan peneliti terhadap penelitian terdahulu:

Tabel 2. 2 Posisi Penelitian

No	Peneliti	Metode			Luaran			Klasifikasi Penyakit
		<i>Support Vector Machine (SVM)</i>	<i>Naïve Bayes Classifier (NBC)</i>	<i>Maximum Entropy (MAXENT)</i>	Perbandingan Akurasi Algoritma	Opini Publik/Masyarakat (Positif/Negatif) di <i>Twitter</i> Terhadap Sebuah Objek	Opini Publik/Masyarakat (Positif/Negatif) di <i>Twitter</i> Terhadap Sebuah Subjek	
1	(Putranti & Winarko, 2014)	√		√	√	√		
2	(Fathan Hidayatullah & Sn, 2014)	√	√		√		√	
3	(Novantirani et al., 2015)	√				√		
4	(Buntoro, 2016)	√	√		√	√		
5	(Faradhillah et al., 2016)	√				√		
6	(Pintoko & Lhaksana, 2018)		√			√		

No	Peneliti	Metode			Luaran			Klasifikasi Penyakit
		<i>Support Vector Machine (SVM)</i>	<i>Naïve Bayes Classifier (NBC)</i>	<i>Maximum Entropy (MAXENT)</i>	Perbandingan Akurasi Algoritma	Opini Publik/Masyarakat (Positif/Negatif) di <i>Twitter</i> Terhadap Sebuah Objek	Opini Publik/Masyarakat (Positif/Negatif) di <i>Twitter</i> Terhadap Sebuah Subjek	
7	(Prakoso et al., 2017)	√		√	√	√		
8	(Rofiqoh et al., 2017)	√				√		
9	(Fanani, 2017)	√	√			√	√	
10	(Aprilla et al., 2018)	√	√	√	√			√
11	(Peneliti, 2021)	√				√		



Gambar 2. 1 Perbandingan Akurasi Penelitian Terdahulu

Berdasarkan kajian empiris dapat diketahui bahwa menggunakan algoritma *Support Vector Machine* hasil akurasi yang didapat cukup tinggi, dan belum ditemukan penelitian yang meneliti tentang persepsi masyarakat tentang vaksin covid-19. Maka dari itu diharapkan penelitian ini dapat dijadikan gambaran untuk mengetahui faktor apa saja yang menjadi penyebab masyarakat masih menolak vaksin covid-19.



2.2 Kajian Deduktif

2.2.1 Imunisasi

Imunisasi berasal dari kata imun yang berartiresisten atau kebal. Imunisasi merupakan suatu program yang dengan sengaja memasukan antigen lemah agar merangsang antibodi keluar sehingga memasukan antigen lemah agar merangsang antibodi keluar sehingga tubuh dapat resissten terhadap penyakit tertentu. Sistem imun tubuh mempunyai suatu sistem memori (daya ingat). Ketika vaksin masuk ke dalam tubuh, maka akan dibentuk antibodi untuk melawan vaksin tersebut dan sistem memori akan menyimpannya sebagai suatu pengalaman (Proverawati & Andhini, 2010).

Imunisasi dapat dilakukan pada orang dewasa ataupun anak-anak, pada anak-anak, karena sistem imun yang belum sempurna. Sedangkan pada usia 60 tahun terjadi penurunan sistem nonspesifik seperti produksi air mata menurun, mekanisme batuk tidak efektif, gangguan pengaturan suhu, dan perubahan sel sistem imun, baik seluler maupun hormonal. Dengan demikian usia lanjut lebih rentan terhadap infeksi, namun usia lanjut masih menunjukkan respon yang baik terhadap polisakarida bakteri, sehingga pemberian vaksin dapat meningkatkan antibodi dengan efektif (Proverawati & Andhini, 2010).

2.2.2 Vaksin

Vaksin berasal dari istilah kata latin *variola* *vaksin* *cinae* (cacar sapi). Pada tahun 1798 Edward Jenner menunjukkan bahwa vaksin tersebut bisa mencegah cacar pada manusia. Vaksin merupakan produk biologis yang dihasilkan dari mikroorganismen hidup, dan meningkatkan kekebalan terhadap penyakit baik mencegah atau mengobati penyakit. Vaksin diberikan dalam bentuk cari, baik suntikan bahkan

melalui mulut (Plotkin, 2013). Menurut FI edisi V, vaksin adalah sediaan yang mengandung zat anti genik yang mampu menimbulkan kekebalan aktif pada manusia. Vaksin dibuat dari bakteri, riketsia dan dapat berupa suspense organisme hidup atau inaktif atau fraksi-fraksinya atau toksoid (Sunarti, 2012).

Vaksinasi berasal dari kata “*vaccine*” yaitu zat yang dapat merangsang timbulnya kekebalan aktif seperti BCG, Polio, DPT, Hepatitis B, dan lain-lain. Vaksin juga menghasilkan sistem kekebalan tubuh terhadap penyakit atau virus baik secara oral maupun injeksi (Kemenkes, 2015).

2.2.3 Twitter

Twitter merupakan salah satu media jejaring sosial yang dapat digunakan dengan mendaftarkan diri sebagai pengguna *Twitter* melalui website atau aplikasi. Pengguna *Twitter* dapat membuat *tweet* dengan kapasitas 140 karakter termasuk spasi dan tanda baca, tetapi *Twitter* telah memperbanyak jumlah *tweet* menjadi 280 karakter pada bulan September 2017 (Maulana, 2017).

Twitter menyediakan *Application Programming Interface* (API) merupakan suatu program atau aplikasi yang diciptakan oleh suatu perusahaan tertentu dengan tujuan mempermudah pihak aplikasi lain dalam mengakses aplikasi tersebut. *Twitter* API diciptakan untuk mempermudah pihak-pihak lain yang ingin mengambil informasi atau data dari *Twitter*. Sebelum menggunakan *Twitter* API, pengguna harus memiliki *customer key* dan *customer secret* dengan tujuan agar aplikasi *web* yang dibentuk dapat diketahui oleh pihak *Twitter* (Rustian, 2019).

Berikut merupakan beberapa fitur yang terdapat dalam *Twitter* (Pakar Online Indonesia, 2012):

1. *Following*

Merupakan akun pengguna *Twitter* yang diikuti akun pengguna *Twitter* lainnya.

2. *Followers*

Merupakan akun pengguna *Twitter* yang mengikuti akun pengguna *Twitter* lainnya.

3. *Tweet*

Pesan yang terdapat pada *Twitter* dengan kapasitas 280 karakter.

4. *Retweet*

Tweet yang telah dibagikan dan dibagikan ulang oleh pengguna lainnya

5. *Mention*

Melibatkan beberapa pengguna pada pesan yang dibagikan dengan awalan “@” pada *username* mereka.

6. *Trending Topics*

Merupakan sepuluh topik yang sedang ramai dibicarakan di *twitter* pada waktu tertentu.

7. *Twitter Search*

Fasilitas yang diberikan agar pengguna dapat lebih mudah mencari subjek, nama, tempat atau kata tertentu.

8. *Direct Message*

Digunakan untuk mengirimkan pesan pribadi ke pengguna lainnya yang bersifat privasi.

9. *Hastag*

Merupakan tanda “#” yang digunakan untuk menglompokan konten.

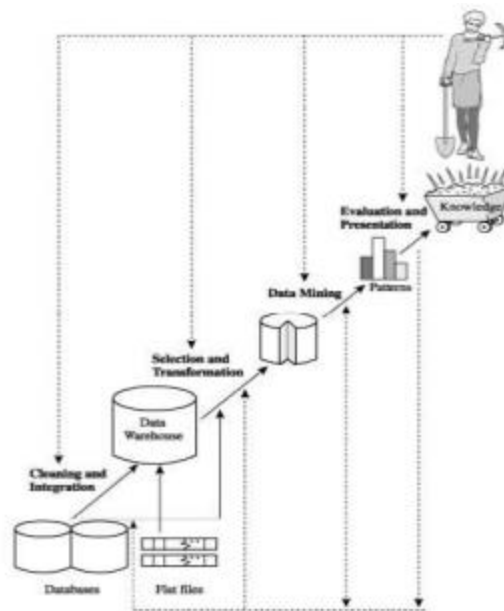
2.2.4 *Text Mining*

Text mining adalah suatu langkah dari analisis teks yang secara otomatis dilakukan oleh komputer dengan tujuan menggali informasi yang berkualitas dari suatu rangkaian teks yang terangkum dalam sebuah dokumen (Han & Kamber, 2006). *Text Mining* merupakan salah satu proses yang bidang dipakai untuk mengerjakan klasifikasi dokumen yang berusaha untuk menemukan bentuk yang menarik dari sekumpulan data yang terkumpul. *Text Mining* merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah mengerjakan ekstraksi pengetahuan

dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks (Samuel & Antonius, 2015). *Text Mining* dalam melakukan analisis, data yang digunakan adalah data yang diperoleh dari dokumen yang bertujuan untuk mencari kata yang dapat menggantikan dari dokumen kata-kata yang lain. *Text Mining* memiliki define sebagai sebuah proses penemuan informasi, relasi dan fakta yang tersembunyi di dalam teks ketika dilakukan pemrosesan dan analisis data dalam jumlah besar, struktur teks yang kompleks dan tidak lengkap, dimensi tinggi, serta data yang *noise*. *Text Mining* memiliki berbagai kegunaan yang dibutuhkan dalam kehidupan sehari-hari (Adiwijaya, 2006).

2.2.5 Data Mining

Data *Mining* merupakan sebuah proses untuk mendapatkan informasi yang berguna dari kumpulan basis data (Tan, 2006). Data *mining* umumnya digunakan untuk menemukan pengetahuan atau informasi yang tersembunyi dalam basis data. Secara umum, data *mining* merupakan proses yang menggunakan teknik matematik, statistic, dan *machine learning* untuk mengidentifikasi dan mengekstrakan informasi dan berguna yang tersimpan dalam basis data Turban et all., (2005). *Knowledge Discovery in Database* (KDD) merupakan sebutan lain dari data mining yang merupakan suatu proses pengambilan informasi yang bermanfaat dan tidak diketahui sebelumnya dari kumpulan sebuah data (Bramer, 2007).



Gambar 2. 2 Tahapan dalam Knowledge Discovery From Database (KDD)

(Han & Kamber, 2006)

Berdasarkan KDD pada Gambar 3.1 terdapat urutan proses berikut:

- a. Pembersihan Data (*Data Cleaning*)
Pembersihan data digunakan untuk menghilangkan *noise* dan data yang tidak konsisten. Penghapusan data yang tidak memiliki kelengkapan atribut sesuai yang dibutuhkan dilakukan pada tahap ini.
- b. Integrasi Data (*Data Integration*)
Merupakan proses penggabungan data dari beberapa sumber data untuk diolah.
- c. Seleksi Data (*Data Selection*)
Proses pengambilan data yang berkaitan dengan analisis yang akan digunakan dalam proses data *mining*.
- d. Tranformasi Data (*Data Transformation*)
Suatu proses transformasi data kedalam bentuk yang diinginkan dalam *mining*.
- e. Penambangan Data (*Data Mining*)
Merupakan proses yang penting dalam KDD yang melibatkan teknik tertentu untuk memperoleh suatu pola dari data yang digunakan.

f. Evaluasi Pola (*Pattern Evaluation*)

Merupakan proses untuk mengkaji kebenaran dari suatu pola data yang mewakili *knowledge* pada data.

g. Representasi Pengetahuan (*Knowledge Representation*)

Suatu proses representasi secara visual kepada pengguna dalam mempermudah pemahaman mengenai hasil dari data *mining*.

2.2.6 *Machine Learning*

Machine Learning atau disebut juga pembelajaran mesin yang merupakan pendekatan dari kecerdasan buatan (*artificial intelegent*) yang banyak digunakan untuk menirukan dan menggantikan perilaku manusia untuk menyelesaikan masalah (Ahmad, 2017). *Machine Learning* adalah suatu proses dalam kecerdasan buatan yang memiliki hubungan terhadap proses pembelajaran dan pemrograman dengan menggunakan data masa lampau. *Machine learning* bekerja apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data besar sehingga menemukan pola tertentu. Data merupakan bahan input yang akan digunakan dalam melakukan pembelajaran atau *training* terdapat dua data yaitu data *testing* dan data *training*, dimana data *training* berfungsi untuk melatih algoritma dalam *machine learning* dan data *testing* berfungsi untuk mengetahui performa dari algoritma dalam *machine learning* yang telah dilatih yaitu ketika menemukan data baru yang belum pernah diberikan dalam data *training*.

Machine learning menggunakan teknik untuk menangani data besar (*big data*) dengan cara yang cerdas untuk memberikan hasil yang tepat. Berdasarkan teknik pembelajarannya, tipe – tipe *machine learning* dapat dibedakan menjadi *supervised learning*, *unsupervised learning*, semi *supervised learning* dan *reinforcement learning*.

2.2.7 *Pre-Processing*

Pada tahap *preprocessing* dilakukan agar teks dapat menjadi data yang dapat diolah lebih lanjut (Falahah, & Nur, D.D, 2015). Tujuan dilakukannya *preprocessing* yaitu mengubah informasi dari tiap-tiap sumber data ke dalam bentuk atau format yang baku sebelum menerapkan berbagai metode-metode pengambilan data terhadap dokumen yang akan diproses (Feldman & Sanger, 2006). Adapun tahapan *preprocessing* yaitu:

a. ***Cleaning***

Cleaning merupakan tahap yang dilakukan untuk membersihkan dokumen dari kata-kata yang tidak diperlukan dan tidak dibutuhkan untuk mengurangi *noisi*. Selain pembersihan kata, dilakukan juga penghilangan tanda baca seperti titik (.), koma (,), dan tanda baca lainnya (~! @\$%^&*()_- +=[]{};:”>, <). Selain itu dilakukan pengubahan huruf dalam dokumen menjadi huruf kecil semua (*lowercase*). Pengubahan menjadi huruf kecil ini biasa disebut juga *casefolding*.

b. ***Filtering***

Filtering atau biasa juga disebut *stopward removal* merupakan penghapusan kata-kata yang kurang bermakna atau kurang penting seperti “dan”, “atau”, “saya”, dan lain sebagainya (Putri D. , 2016).

c. ***Tokenizing***

Tokenizing merupakan tahap pemotongan kata-atau potongan kalimat yang biasa disebut token. Kata dalam dokumen yang dipisahkan adalah kata pada kalimat yang terpisah oleh spasi.

d. ***Stemming***

Stemming merupakan tahap mengubah kata menjadi kata dasar (*roof*) dari setiap kata hasil proses *filtering* yang mengandung imbuhan.

2.2.8 Sentiment Analysis

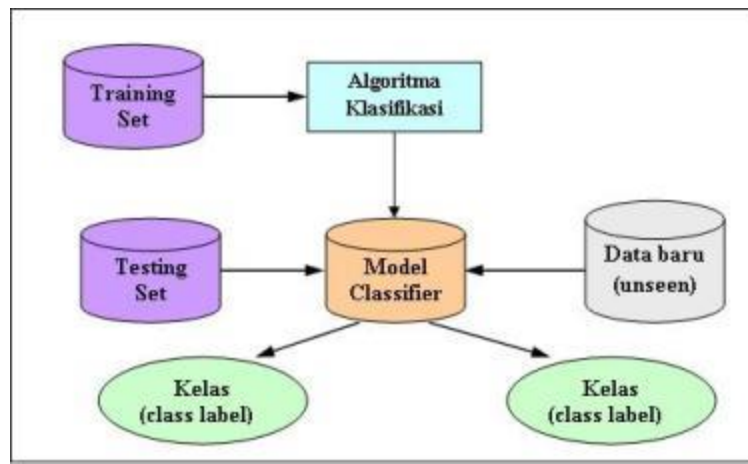
Analisis sentiment atau *opinion mining* merupakan studi komputasi mengenai pendapat, perilaku, dan emosi seseorang terhadap entitas tersebut dapat mewakili individu, kejadian atau topik (Kristiyanti, 2015). Melalui analisis sentiment dapat diketahui opini public terkait suatu hal baik berupa produk, jasa, aplikasi, film, musik, dan lain sebagainya melalui review pengguna pada berbagai situs. Analisis sentiment digunakan untuk melihat kecenderungan apakah pendapat atau opini pada suatu masalah cenderung termasuk dalam opini positif atau opini negatif (Rozi, 2012).

2.2.9 Classification

Klasifikasi merupakan teknik analisis data yang digunakan untuk menghasilkan model prediksi untuk mendeskripsikan label atau kelas data (Han, J., & Kamber, M., 2006). Sedangkan menurut Prasetyo (2012) klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Klasifikasi tergolong dalam *supervised method*.

Menurut Han, J., & Kamber, M., (2006) *classification* memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian *data class* yang disebut *learned model*. Model tersebut dibangun dengan menganalisa *record database*. Setiap *record* diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang disebut *class label* atribut. Akibat terdapat *class label* maka tahap ini juga dikenal dengan *supervised learning*. Berbeda dengan *unsupervised learning* atau dikenal dengan *clustering* yang tidak memerlukan *class label*. Tahap pertama ini juga disebut sebagai tahap pembelajaran. Sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis data training. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $y = f(x)$ dimana y adalah kelas hasil prediksi dan x adalah record yang ingin diprediksi class-nya.

Berikut merupakan bagian proses klasifikasi data sampel menggunakan model *classifier* untuk mendapatkan hasil prediksi:



Gambar 2. 3 Klasifikasi Menggunakan *Model Classifier* (KDD)

Beberapa persiapan yang dilakukan untuk mendapatkan hasil klasifikasi yang baik di antaranya adalah (Han, J., & Kamber, M., 2006):

1. Pembersihan Data

Pembersihan data ini dilakukan untuk mengurangi kecacatan data di dalam data pelatihan, beberapa metode yang digunakan diantaranya dengan teknik *smoothing* untuk menghilangkan *noise data*, melengkapi data yang hilang dan sebagainya.

2. Analisis Relevansi

Dari beberapa atribut yang akan digunakan untuk proses klasifikasi mungkin saja terdapat atribut yang sangat berhubungan kuat satu sama lain, kedua atribut ini memiliki kemiripan sehingga menyebabkan proses klasifikasi menjadi tidak optimal, maka salah satu dari atribut ini dapat dibuang.

Hasil klasifikasi dan prediksi dapat dievaluasi menggunakan beberapa kriteria (Han, J., & Kamber, M., 2006)

1. Akurasi

Akurasi digunakan untuk mengetahui kemampuan model klasifikasi untuk dapat memberikan ketepatan hasil prediksi.

2. Kecepatan

Mengetahui kecepatan iterasi untuk mendapatkan model klasifikasi dan iterasi mendapatkan hasil prediksi.

2.2.10 Ukuran Evaluasi Model Klasifikasi

Proses evaluasi dilakukan dengan menghitung suatu ukuran tertentu terhadap himpunan data uji, yakni data yang tidak digunakan dalam proses pembuatan model klasifikasi tersebut. *Confusion matrix* merupakan matrix yang berisi informasi mengenai klasifikasi aktual yang akan diprediksi oleh sistem klasifikasi (Kohavi, R., & Provost, F., 1998). Sistem klasifikasi dibentuk dari pemetaan suatu baris data dan *output* suatu hasil prediksi kelas dari data tersebut. Pada suatu klasifikasi baris data dapat menghasilkan empat kemungkinan yang digunakan untuk menilai dan mengevaluasi proses klasifikasi. Apabila data positif dan tepat diprediksi positif maka disebut *true positive*, namun jika salah dan terprediksi negatif maka disebut *false negative*. Apabila data negatif dan tepat diprediksi negatif maka disebut *true negative*, namun jika salah dan terprediksi positif maka disebut *false positive* (Fawcett, 2006).

Tabel 2. 3 Model Klasifikasi

	Kelas	Positif	Negatif
Prediksi	Positif	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	Negatif	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Confusion matrix sangat berguna untuk menganalisis kualitas *classifier* dalam mengenali tuple-tuple dari kelas yang ada. TP dan TN menyatakan bahwa *classifier* mengenali tuple dengan benar, artinya tuple positif dikenali sebagai positif dan tuple negative dikenali sebagai negative. Sebaliknya, FP dan FN menyatakan bahwa *classifier* salah dalam mengenali tuple, *tuple negative* dikenali sebagai positif dan *tuple positive* dikenali sebagai negative. Terdapat beberapa rumus umum yang dapat

digunakan untuk menghitung performa klasifikasi. Hasil dari nilai akurasi, presisi, dan *recall* biasa ditampilkan dalam presentase.

a. *Accuracy*

Akurasi adalah jumlah proporsi prediksi yang benar.

b. *Precision*

Precision adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen yang terpilih oleh sistem.

c. *Recall*

Recall adalah proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks relevan yang ada pada koleksi.

2.2.11 Fitur dan Pembobotan

Pembobotan merupakan metode untuk mengubah input data menjadi suatu fitur vector. Metode pembobotan yang umum digunakan adalah *bag-of-feature*. Sebagai contoh terdapat sederetan fitur seperti pada vector $\{f_1, f_2, \dots, f_n\}$ dimana yaitu sekumpulan fitur-fitur sebanyak n yang sudah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vektor dari data adalah vektor.

a. *Term Presence*

Term Presence (TP) ialah metode pembobotan pada suatu dokumen teks yang melihat keberadaan daftar kata-kata (*term*) atau fitur yang ada pada corpus terhadap suatu dokumen. Jika suatu fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang sedang diboboti maka nilai fitur tersebut pada *feature vector* akan diberi nilai 1 dan tidak menghiraukan jumlah kemunculan fitur tersebut. Jika fitur tersebut tidak ada pada dokumen maka diberi nilai 0 pada *feature space* (O’Keefe, & Koprinska., 2009).

b. *Term Frequency*

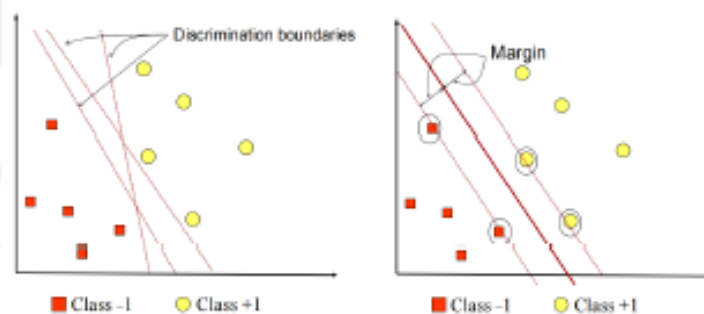
Term Frequency (TF) memiliki kesamaan dengan TP yang telah dijelaskan sebelumnya, tapi yang membedakan adalah TF menghitung jumlah kemunculan fitur acuan pada suatu dokumen bukan hanya keberadaan fitur tersebut (O’Keefe, & Koprinska., 2009).

c. *Term Frequency -Inverse Document Frequency* (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) merupakan algoritma pembobotan tersusun dari dua nilai yang berasal dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (TF-IDF). *Output* dengan fitur/*term* tersebut dengan kata yang sering muncul pada dokumen akan menghasilkan nilai TF-IDF yang tinggi. Sementara, fitur yang sering muncul pada dokumen akan menghasilkan nilai rendah (O’Keefe, & Koprinska., 2009).

2.2.12 Support Vector Machine

Menurut Prasetyo (2012), metode SVM (*Support Vector Machine*) merupakan teori pembelajaran statistic dan dapat memberikan hasil yang lebih baik dari pada metode yang lain. SVM dapat bekerja dengan baik pada data dengan berdimensi set tinggi. Selain itu, SVM menggunakan teknik kernel dan hanya sejumlah data yang terpilih yang berkontribusi untuk membangun model klasifikasi. Hal tersebut menjadi kelebihan SVM, karena tidak semua data latih akan dilihat untuk dilibatkan dalam setiap literasi pelatihannya. Konsep dasar dari SVM yaitu mencari *hyperplane* (*maximal margin hyperplane*) terbaik yang memiliki fungsi untuk pemisah dua buah kelas pada input.



Gambar 2. 4 Klasifikasi Menggunakan Model Classifier (KDD)

Sumber: (Nugroho et al., 2003)

Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin batasan dari *hyperplane* itu sendiri dan mencari titik jarak maksimalnya. Margin batasan adalah jarak antara *hyperplane* tersebut dengan *pattern* (atau disebut

dengan *support vector*) terdekat dari masing-masing kelas. Garis *solid* pada gambar 2.3 menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector* (Nugroho et al., 2003)

2.2.13 Asosiasi Kata

Istilah korelasi sering digunakan untuk menyatakan hubungan dua atau lebih variabel sifatnya kuantitatif. Sedangkan istilah asosiasi sering dimaknai keeratan hubungan antara dua atau lebih variabel yang sifatnya kualitatif (Ulwan, 2016). Tujuan analisis korelasi adalah untuk mencari hubungan variabel bebas (X) dengan variabel terikat (Y), dengan ketentuan data memiliki syarat-syarat tertentu (Fadlisyah., 2014).

Asosiasi merupakan proses pencarian hubungan antar elemen data. Pada penelitian ini digunakan pendekatan asosiasi untuk menemukan hubungan antar ulasan pada masing-masing klasifikasi ulasan positif dan ulasan negative pengguna sehingga mendapatkan informasi yang dapat dijadikan bahan rujukan dalam pengembangan serta untuk meningkatkan kualitas perusahaan.

2.2.14 Diagram Pohon

Pohon masalah (problem tree) merupakan sebuah pendekatan/ metode yang digunakan untuk identifikasi penyebab suatu masalah. Analisis pohon masalah dilakukan dengan membentuk pola pikir yang lebih terstruktur mengenai komponen sebab akibat yang berkaitan dengan masalah yang telah diprioritaskan. Metode ini dapat diterapkan apabila sudah dilakukan identifikasi dan penentuan prioritas masalah. Menurut Silverman & Lori L (1994) menggunakan istilah *tree diagram* dan menyatakan diagram sistematis atau diagram pohon dirancang untuk mengurutkan hubungan sebab akibat.

BAB III

METODE PENELITIAN

3.1 Objek dan Subjek Penelitian

Objek penelitian ini adalah salah satu *Hot Topics* di Indonesia saat ini yaitu vaksinasi Covid-19. Adapun subjek penelitian ini adalah persepsi masyarakat terhadap vaksin Covid-19 melalui *review* pada *database website Twitter*.

3.2 Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah semua data *review* atau ulasan terhadap *tweet* mengenai vaksin Covid-19 dari *database website Twitter*. Sedangkan sampel diambil dari suatu populasi yang mana pada penelitian ini digunakan sampel ulasan pengguna *Twitter* dengan rentang waktu dari bulan Januari 2021 sampai dengan bulan Februari 2021. Dengan jumlah sampel sebanyak 3131 ulasan.

3.3 Pengumpulan Data

3.3.1 Metode Pengumpulan Data

Metode yang digunakan untuk proses pengumpulan data dari penelitian ini adalah sebagai berikut:

1. *Twitter API*

Data pada penelitian ini adalah data sekunder yang diambil secara online dengan bantuan *Twitter API* dengan cara mendaftar pada *Twitter Developer*. Melalui ini didapatkan informasi dari *website* secara otomatis tanpa perlu menyalin secara manual.

2. Studi Pustaka

Studi pustaka dilakukan dengan menggunakan beberapa kajian literatur, buku, maupun referensi jurnal yang sekiranya berkaitan dengan tujuan, rumusan, batasan, dan metode penelitian

3.3.2 Jenis Data

Jenis data yang digunakan pada penelitian ini adalah data sekunder, hal ini dikarenakan peneliti mendapatkan data melalui media perantara. Pada penelitian ini data sekunder berperan sebagai *main data* dalam proses pengolahan untuk mengetahui persepsi masyarakat terhadap vaksin Covid-19. Adapun sumber data yang digunakan dalam penelitian ini adalah *dataset reviews* terhadap *topics* tersebut yang diambil dari *website Twitter*.

3.3.3 Variabel dan Definisi Operasional Variabel

Adapun variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 3. 1 Variabel Penelitian

Variabel	Definisi Operasional Variabel
Teks	Teks yang dimaksud merupakan tanggapan mengenai vaksin Covid-19 yang terdapat pada media sosial <i>Twitter</i> .

Label	Label yang dimaksud merupakan kategori dari tanggapan yang telah diperoleh. Peneliti mengkategorikan tanggapan ke dalam dua kelas, yaitu tanggapan positif dan tanggapan negatif.
Waktu	Waktu berupa tanggal, bulan, dan tahun. saat yang ada pada <i>tweet</i> tersebut. Variabel waktu perlu diperhatikan mengingat ulasan yang diambil adalah berupa sampel dengan waktu tertentu yang telah ditentukan sebelumnya.

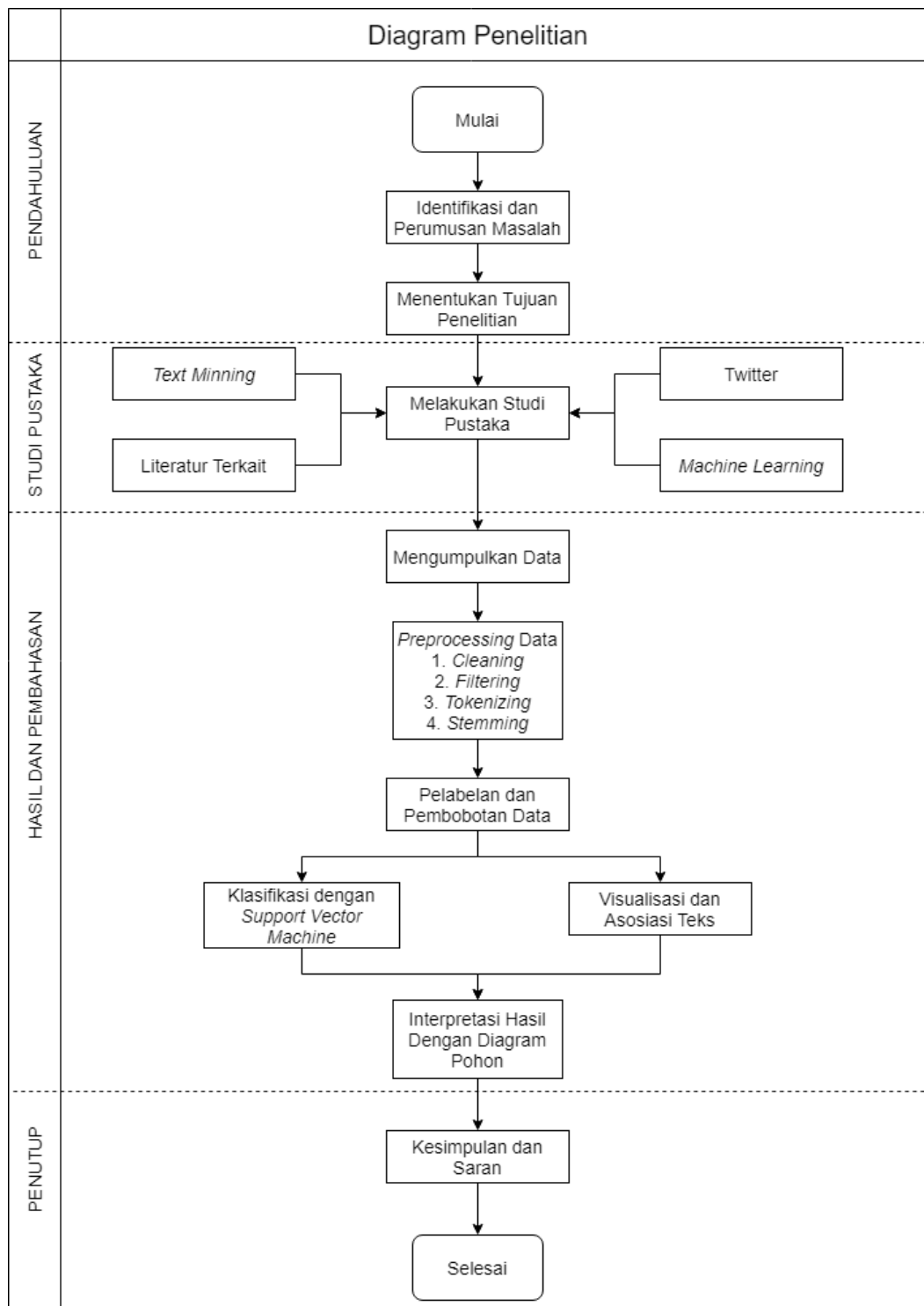
3.4 Metode Analisis Data

Berikut ini merupakan metode analisis data yang digunakan dalam penelitian ini, antara lain:

1. Metode *Machine Learning* yang digunakan adalah *Support Vector Machine* yang digunakan untuk melakukan klasifikasi ulasan yang berbentuk kelas positif maupun negatif.
2. *Wordcloud*, digunakan untuk melakukan visualisasi kata yang paling banyak digunakan atau paling sering muncul pada *tweet* terkait topics tersebut.
3. *Association*, digunakan untuk mengidentifikasi dan membentuk pola kata yang berasosiasi dengan kata lainnya untuk mendapatkan informasi yang dianggap penting.
4. Diagram pohon (*Tree diagram*), menemukan akar permasalahan dan untuk mendapatkan solusi terbaik.

3.5 Alur Penelitian

Berikut merupakan alur dari penelitian ini:



Gambar 3. 1 Alur Penelitian

Berdasarkan diagram alir penelitian, langkah-langkah pada penelitian ini terdiri atas sebagai berikut:

1. Pendahuluan

Tahapan ini merupakan tahap awal dalam melakukan penelitian yang meliputi identifikasi dan perumusan masalah, dan menentukan tujuan penelitian. Identifikasi masalah dilakukan dengan melihat kondisi yang ada sekarang seperti saat ini yang menjadi *hot topics* yaitu vaksin Covid-19. Kemudian menetapkan tujuan penelitian berdasarkan identifikasi masalah yang dibuat.

2. Studi Pustaka

Untuk dapat mengetahui metode yang sekiranya tepat untuk penelitian ini, maka dilakukan studi pustaka yang berkaitan dengan topik yang telah ditentukan yaitu analisis sentimen. Berikut ini beberapa studi pustaka yang menjadi referensi:

- a. *Text Mining*
- b. *Machine Learning*
- c. Literatur terkait
- d. *Twitter*

3. Hasil dan Pembahasan

Pada tahapan ini terdapat pengumpulan data, *preprocessing data*, dan pelabelan dan pembobotan data. Pengumpulan data dilakukan dengan aplikasi Rstudio dan mendapatkan code akses dari *Twitter Developer* yaitu *Twitter API*. *Preprocessing data* digunakan untuk menyeleksi data dan mengubahnya menjadi data yang terstruktur dan merupakan tahapan awal untuk mengubah struktur isi data untuk menjadi format yang sesuai agar dapat diproses oleh algoritma *Support Vector Machine*. Proses dari *preprocessing* ini dilakukan menggunakan *software R*. Kemudian yang terakhir pelabelan dan pembobotan data digunakan untuk melakukan

klasifikasi ulasan yang berbentuk kelas positif maupun negatif, dan mengekstraksi keseluruhan informasi sehingga dapat diketahui kata yang paling sering digunakan pada komentar yang ada. Oleh karena ini bentuk dari visualisasi dan asosiasi ini nantinya akan berbentuk *wordcloud* dan *barplot*. Setelah mengetahui visualisasi dari sentimen negatif, langkah selanjutnya adalah membuat diagram pohon untuk menganalisis faktor-faktor yang menjadi masalah di mata masyarakat terhadap vaksin Covid-19 dan mengidentifikasi hal apa saja yang perlu dilakukan dalam rangka menanggulangi permasalahan tersebut

4. Penutup

Tahap terakhir dari penelitian ini adalah menarik kesimpulan dan saran berdasarkan dari hasil penelitian.

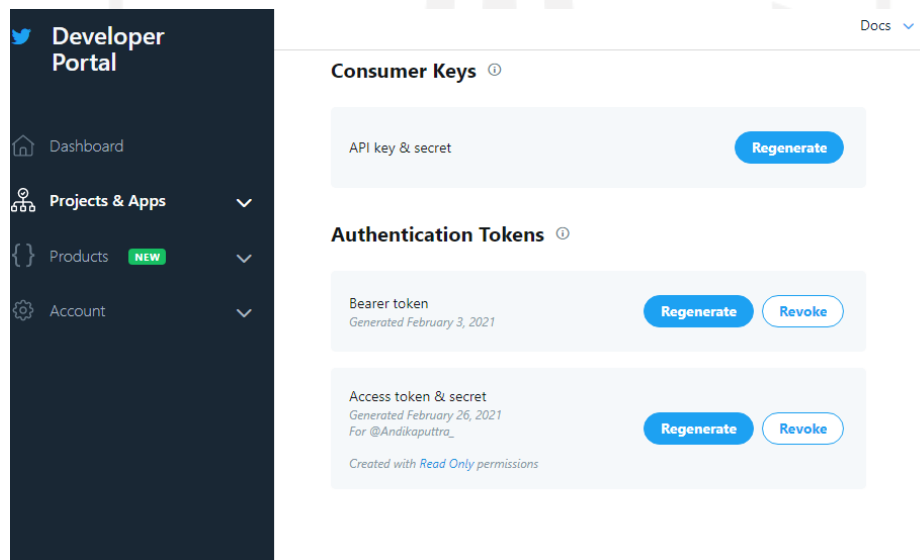


BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Pada penelitian ini dilakukan pengambilan data berupa persepsi masyarakat terhadap vaksin Covid-19 melalui *review* pada *database website Twitter*. Pada *Twitter* dibutuhkan sebuah kode didapat dari *Twitter API*. *Twitter API* merupakan aplikasi yang diciptakan oleh pihak *twitter* dengan tujuan agar mempermudah pihak developer untuk mengakses informasi *web Twitter*. Pendaftaran *API* digunakan untuk mengonfirmasi kepada pihak *Twitter* agar memberikan izin menjelajahi lebih luas terkait dengan data yang berkaitan dengan *Twitter*.



Gambar 4. 1 Twitter Developer

Setelah registrasi dan bergabung dengan *Twitter API*, kemudian hasil dari registrasi tersebut didapatkan beberapa kode berupa *consumer key*, *consumer secret*, *access token*, dan *access key* dari *Twitter*. Kode *API* tersebut adalah sebagai jembatan antara *Twitter* dengan aplikasi lainnya, dalam penelitian ini kode tersebut dapat digunakan untuk proses integrasi antara *Twitter API* dengan *R studio*.

4.2 Pengolahan Data

Data yang telah didapat kemudian diolah agar didapatkan informasi dari data tersebut. Pengolahan data dilakukan pada masyarakat terhadap vaksin Covid-19 dengan sampel 3131. Terdapat dua pengolahan data yang dilakukan yang terdiri dari analisis deskriptif dan juga *pre-processing* data, yaitu sebagai berikut

4.2.1 *Pre-processing* Data

Data dalam penelitian ini merupakan data *tweet* tanggapan yang diambil dari media sosial *Twitter* yang memiliki berbagai macam gaya penulisan sehingga data yang diperoleh merupakan data yang tidak terstruktur. Pada tahap ini bertujuan untuk membersihkan data-data dari *noise* dan membenarkan bahasa seperti menghilangkan singkatan, bahasa gaul, serta menghapus kata yang tidak diperlukan, karena data awal yang didapatkan berupa data yang tidak terstruktur maka dilakukan tahap *preprocessing* agar data tersebut dapat di analisis. Berikut merupakan contoh data yang akan dilakukan proses *Pre-processing*:

Tabel 4. 1 Contoh Data Untuk *Preprocessing*

No	Data
1	Anjaaayyy mau dong divaksinFb kk #VaksinUntukRakyatIndonesia https://t.co/KUlgGNwYEG
2	Yuk Patuhi 3M #VaksinUntukRakyatIndonesia https://t.co/62O6mz20ww

3 DENGARIN INI..KATANYA VAKSIN SINOVAC DARI CINA ITU RONGSOKAN..ADA YG BERANI TANGKAP IBU INI NGK..RAKYAT JANGAN MAU DIKIBULIN PEMERINTAH..

Dalam tabel diatas menunjukkan beberapa contoh *tweet* terkait sikap komentar masyarakat mengenai vaksin Covid-19. Berikut merupakan tahapan-tahapan *preprocessing*:

1. *Cleaning Data*

Hasil dari *crawling* merupakan data mentah atau data yang diperoleh masih terdapat unsur simbol, *URL*, dan sebagainya yang tidak mempunyai arti pada kalimat tersebut. Proses *cleaning* bertujuan untuk merapikan dan membersihkan kalimat dari kata-kata yang tidak memiliki arti sehingga lebih mudah dan cepat dalam mendapatkan informasi dari data yang didapat. Berikut merupakan contoh kalimat ulasan hasil dari *cleaning* data:

- Menghapus URL:

Tabel 4. 2 Menghapus URL Pada Data

Data Input	Data Output
Anjaaayyy mau dong divaksinFb kk #VaksinUntukRakyatIndonesia https://t.co/KUlgGNwYEg	Anjaaayyy mau dong divaksinFb kk #VaksinUntukRakyatIndonesia

- Menghapus tanda baca

Tabel 4. 3 Menghapus Tanda Baca Pada Data

Data Input	Data Output
Yuk Patuhi 3M #VaksinUntukRakyatIndonesia https://t.co/62O6mz20ww	Yuk Patuhi 3M VaksinUntukRakyatIndonesia https://t.co/62O6mz20ww

- Menyeragamkan huruf menjadi huruf kecil (*lower-case*)

Tabel 4. 4 Menyeragamkan Huruf Menjadi Huruf Kecil (*Lower-Case*)

Data Input	Data Output
DENGARIN INI..KATANYA	dengarin ini..katanya
VAKSIN SINOVAC DARI CINA	sinovac dari cina
ITU RONGSOKAN..ADA YG	itu rongsokan..ada yg
BERANI TANGKAP IBU INI	berani tangkap ibu ini
NGK..RAKYAT JANGAN MAU	ngk..rakyat jangan mau
DIKIBULIN PEMERINTAH..	dikibulin pemerintah..

2. *Tokenizing*

Tokenizing merupakan tahap memisahkan teks menjadi potongan kata yang independen dan tidak saling mempengaruhi atau disebut juga *token*. Melalui *tokenizing* proses perhitungan dari frekuensi kemunculan setiap kata dalam *corpus* dapat lebih mudah terdeteksi. Berikut merupakan contoh penerapan *tokenizing*:

Tabel 4. 5 *Tokenizing*

Data Input	Data Output
Hentikan Penyebarannya, tetap patuhi proses	“hentikan” “penyebaran” “tetap” “patuhi” “proses”

3. *Filtering*

Tahap *filtering* yaitu tahapan untuk mengambil kata-kata yang penting. Proses *filtering* dapat menggunakan algoritma *stopword* (menghapus kata tidak penting). Contoh *stopword* yaitu “yang”, “dan”, “ke”, “dari”, “oleh”, dan lainnya. Kata-kata tersebut merupakan kata yang berfrekuensi tinggi dan dapat ditemukan hampir di setiap kalimat. *Stopword* atau menghapus kata dapat mengurangi ukuran indeks dan waktu pemrosesan serta dapat mengurangi *noise*.

Tabel 4. 6 *Filtering*

Data Input	Data Output
------------	-------------

Mari kita di vaksin	mari vaksin
---------------------	-------------

4. *Stemming*

Stemming merupakan tahap yang dilakukan untuk mengubah kata pada teks menjadi kata dasar. Pada tahap yang dilakukan untuk mengubah kata pada teks menjadi kata dasar. Pada tahap ini dilakukan penghapusan imbuhan pada kata hasil dari proses *filtering*, baik yang mengandung imbuhan induktif maupun deduktif. Pada proses *stemming* tidak diikutsertakan pada proses selanjutnya, hal ini dikarenakan pada proses ini menghilangkan makna-makna penting terkait persepsi masyarakat terhadap vaksin Covid-19. Adapun berikut merupakan contoh proses *stemming* pada ulasan:

Tabel 4. 7 *Stemming*

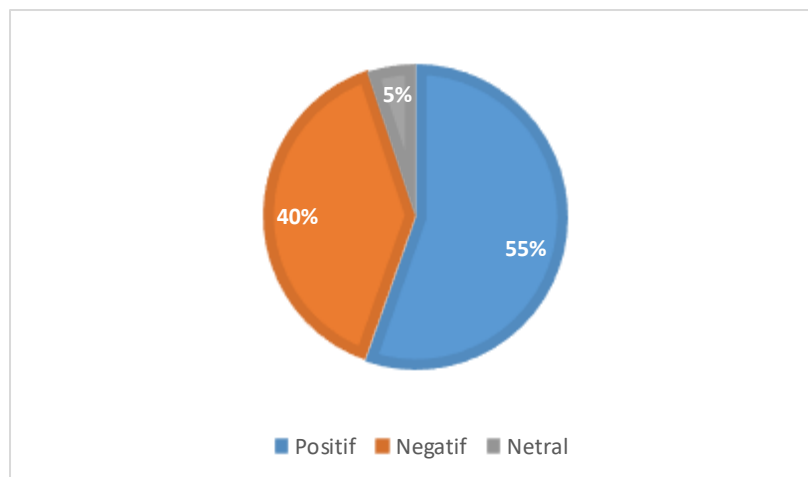
Data Input	Data Output
Tetap menjaga kesehatan	Tetap jaga sehat

4.2.2 Pelabelan dan Pembobotan Kelas Sentimen

Tahap berikutnya setelah mendapatkan skor terhadap sentimen adalah melakukan pelabelan kelas sentimen. Menurut (Buntoro, 2017) pelabelan dilakukan dengan membagi data menjadi tiga kelas sentimen yaitu sentiment positif, sentimen netral, dan sentimen negatif dengan keterangan sebagai berikut:

- Sentimen negatif : skor < 0
- Sentimen netral : skor = 0
- Sentimen positif : skor > 0

Data tanggapan terdiri dari data tanggapan pada bulan Januari dan Februari yang dikategorikan kedalam tiga kelas kategori, yaitu tanggapan positif, negatif, dan netral. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing masing kategori sebagai berikut:



Gambar 4. 2 Jumlah Tanggapan Tentang Vaksin Covid-19

Berdasarkan gambar diatas dapat diketahui bahwa dari total tanggapan yang diperoleh dari bulan Januari sampai bulan Februari yaitu 3131, sebesar 55% atau sebanyak 1735 merupakan tanggapan positif yaitu tanggapan yang mendukung adanya proses vaksinasi Covid-19, sebesar 40% atau sebanyak 1236 merupakan tanggapan negatif yaitu tanggapan yang kurang menyukai suatu hal yang berhubungan dengan vaksin Covid-19 dan sebesar 5% atau sebanyak 160 merupakan tanggapan netral yaitu tanggapan yang tidak mengandung unsur pro dan kontra atau nilainya berimbang. Pada ulasan klasifikasi yang mengandung pernyataan positif seperti membanggakan, ungkapan terimakasih maupun kata-kata pujian dan lain sebagainya dimasukkan ke dalam kategori sentimen positif. Pada ulasan pengklasifikasian yang mengandung pernyataan negatif seperti penghinaan, ketidakpuasan dan sebagainya dimasukkan kedalam kategori negatif. Pada ulasan yang tidak mengandung kata pernyataan positif ataupun negatif serta terdapat ulasan yang mengandung kategori positif dan negatif yang seimbang maka masuk dalam kategori sentimen netra seperti pertanyaan tanpa sentimen, iklan dan sebagainya. Adapun contoh hasil pelabelan data ulasan sebagai berikut:

Tabel 4. 8 Pelabelan Kelas Sentimen Pada Kalimat

Kelas Sentimen	Skor	Ulasan
Positif	3	Terbukti aman halal
Negatif	2	Ragu takut
Netral	0	Kawal proses

Pada penelitian ini dilakukan reduksi kelas dengan mengkategorikan kelas sentimen netral ke dalam kelas sentimen positif atau negatif yang dilakukan secara manual. Apabila kelas sentimen netral tidak teridentifikasi kata sentimen positif ataupun negatif maka akan masuk dalam pengklasifikasian kelas sentimen negatif (Gumilang, 2018). Adapun simulasi perhitungan skor sentimen adalah sebagai berikut:

Tabel 4. 9 Perhitungan Skor Sentimen

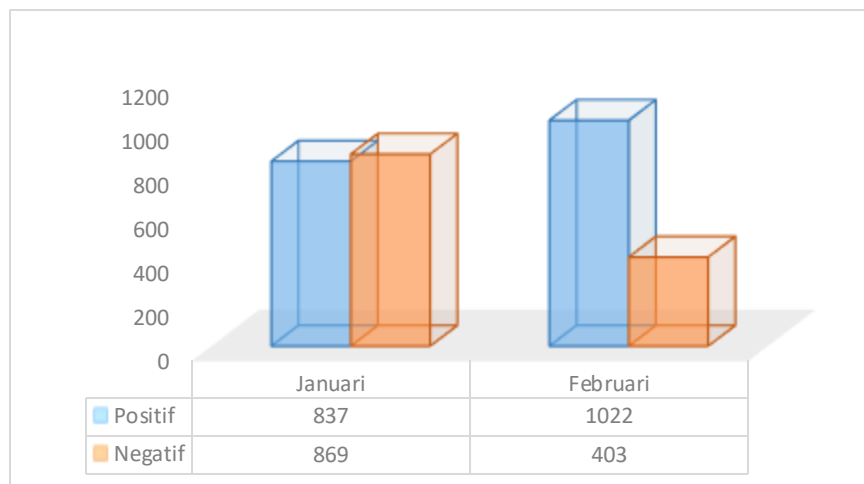
Ulasan	Kata Positif	Kata Negatif
Protokol kesehatan lupa	kesehatan	lupa
Jumlah	1	1

Sehingga didapatkan hasil perhitungan sebagai berikut:

$$\begin{aligned}
 \text{Skor} &= \text{jumlah kata positif} - \text{jumlah kata negatif} \\
 &= 1 - 1 \\
 &= 0
 \end{aligned}$$

Skor akhir yang didapatkan dari simulasi perhitungan bernilai 0, sehingga hasil klasifikasi ulasan adalah negatif.

Setelah kelas netral direduksi, hasil pelabelan kelas sentimen dengan frekuensi terbanyak adalah sentimen positif yaitu sebesar 1859 ulasan sedangkan sentimen negatif sebesar 1272 ulasan. Perbandingan tanggapan yang diperoleh pada merupakan perbandingan dari seluruh data tanggapan yang diperoleh yaitu data tanggapan pada bulan Januari dan Februari. Oleh karena itu, untuk melihat perbedaan perbandingan tanggapan pada masing-masing bulan, peneliti menampilkan perbandingan pada masing-masing bulan, sebagai berikut:



Gambar 4. 3 Tanggapan Tentang Vaksin Covid-19 Pada Bulan Januari dan Februari

Berdasarkan gambar diatas dapat diketahui bahwa pada bulan Januari jumlah tanggapan negatif lebih besar dibandingkan tanggapan positif. Sedangkan pada bulan februari jumlah tanggapan positif lebih besar dibandingkan tanggapan negatif. Pada bulan Januari dari 1705 tanggapan, sebanyak 837 tanggapan merupakan tanggapan positif dan sebanyak 869 tanggapan merupakan tanggapan negatif. Pada bulan Februari dari 1426 tanggapan, sebanyak 1022 tanggapan merupakan tanggapan positif dan sisanya sebanyak 403 tanggapan merupakan tanggapan negatif.

4.2.3 Pembagian Data *Training* dan Data *Testing*

Data *training* digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada, semakin besar data latih yang digunakan, maka akan semakin baik *machine* dalam memahami pola data. Sedangkan data *testing* digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Suthaharan (2015) menyatakan bahwa meskipun penelitian ekstensif belum dilakukan dalam pemilihan rasio yang optimal antara kumpulan data ini, ada beberapa praktik umum dalam memilih ukuran

kumpulan data ini. Berdasarkan *Pareto Principle*, rasio umum yang digunakan adalah 80:20. Adapun pada penelitian ini dilakukan lima kali percobaan dengan perbandingan antara data *training* dan data *testing* yang berbeda, sebagai berikut:

1. Perbandingan data *training* sebesar 50% dan data *testing* 50%

Tabel 4. 10 Perbandingan Data *Training* dan Data *Testing*

Kelas	Jumlah	Data Training (50%)	Data Testing (50%)
Positif	1859	930	929
Negatif	1272	636	636
Total	3131	1566	1565

2. Perbandingan data *training* sebesar 60% dan data *testing* 40%

Tabel 4. 11 Perbandingan Data *Training* dan Data *Testing*

Kelas	Jumlah	Data Training (60%)	Data Testing (40%)
Positif	1859	1115	744
Negatif	1272	763	509
Total	3131	1878	1253

3. Perbandingan data *training* sebesar 70% dan data *testing* 30%

Tabel 4. 12 Perbandingan Data *Training* dan Data *Testing*

Kelas	Jumlah	Data Training (70%)	Data Testing (30%)
Positif	1859	1301	558
Negatif	1272	890	382
Total	3131	2191	940

4. Perbandingan data *training* sebesar 80% dan data *testing* 20%

Tabel 4. 13 Perbandingan Data *Training* dan Data *Testing*

Kelas	Jumlah	Data Training (80%)	Data Testing (20%)
--------------	---------------	----------------------------	---------------------------

Positif	1859	1487	372
Negatif	1272	1018	254
Total	3131	2505	626

5. Perbandingan data *training* sebesar 90% dan data *testing* 10%

Tabel 4. 14 Perbandingan Data *Training* dan Data *Testing*

Kelas	Jumlah	Data Training (90%)	Data Testing (10%)
Positif	1859	1673	186
Negatif	1272	1145	127
Total	3131	2818	313

4.2.4 Klasifikasi dengan *Support Vector Machine*

Proses klasifikasi akan dilakukan dengan cara mempelajari data latih yang telah dibentuk pada sub bab sebelumnya. Pada data latih terdapat review positif dan review negatif dengan perbandingan yang telah ditentukan, kemudian akan dipelajari pola data tersebut dengan algoritma SVM untuk mendeteksi ciri-ciri data pada masing-masing kelas yang telah dibentuk pada data latih. Hasil pelatihan pada algoritma SVM kemudian diuji dengan menggunakan data uji, sehingga didapatkan nilai akurasi dalam memprediksi pada data baru. Proses inilah yang disebut sebagai machine learning.

Pada proses klasifikasi menggunakan algoritma SVM metode yang digunakan adalah kernel untuk memperoleh klasifikasi dengan akurasi terbaik. Metode kernel yang digunakan adalah kernel *Linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Dalam menguji metode kernel yang terbaik menggunakan perbandingan 80:20, karena berdasarkan *Pareto Principle* perbandingan 80:20 adalah rasio umum yang digunakan. Berikut merupakan akurasi tiap metode kernel:

Tabel 4. 15 Perbandingan Akurasi Metode Kernel

Kernel	Akurasi
Linear	91.05%
Polynomial	78.91%
RBF	87.06%
Sigmoid	88.17%

Berdasarkan perbandingan hasil beberapa metode kernel, menunjukkan bahwa kernel *Linear* memiliki tingkat akurasi tertinggi diantara metode kernel yang lainnya. Dengan demikian proses klasifikasi ini akan menggunakan kernel *Linear* dengan algoritma SVM. Adapun tahap-tahap melakukan klasifikasi menggunakan algoritma SVM sebagai berikut:

Tabel 4. 16 Tahap Melakukan Analisis SVM

Script	Fungsi
<pre>setwd("E://KULIAH/METOPEL/SKRIPSI") positiftraining = readLines("positiftraining4.csv") negatiftraining = readLines("negatiftraining4.csv") positiftesting = readLines("positiftesting4.csv") negatiftesting = readLines("negatiftesting4.csv") reviewtraining = c(positiftraining, negatiftraining) reviewtesting= c(positiftesting, negatiftesting) review_all = c(reviewtraining,reviewtesting) sentiment_training = c(rep("positiftraining", length(positiftraining)), rep("negatiftraining", length(negatiftraining))) sentiment_test = c(rep("positiftraining", length(positiftesting)), rep("negatiftraining", length(negatiftesting)))</pre>	<ol style="list-style-type: none"> 1. Mengatur direktori kerja pada program R. 2. Membuka file data latih dan data uji dalam format <i>csv</i>. 3. Mendefinisikan masing-masing data latih dan data uji. 4. Menggabungkan data latih dan data uji yang telah terdefinisi. 5. Mendefinisikan label kelas pada data latih dan data uji. 6. Menggabungkan label kelas dan mengubah tipe data menjadi tipe data faktor.

<i>Script</i>	<i>Fungsi</i>
<pre>sentiment_all = as.factor(c(sentiment_training, sentiment_test)) library(RTextTools) library(e1071) mat = create_matrix(review_all, language = "indonesian",removeStopwords = FALSE, removeNumbers = TRUE, stemWords = FALSE, tm::weightTfIdf) mat = as.matrix(mat) #SVM Linear container<-create_container(mat, sentiment_all, trainSize=1:2191,testSize=2192:3131, virgin=FALSE) models <- train_models(container,"SVM",kernel='linear') results <- classify_models(container, models) table(as.character(sentiment_all[2192:3131]), as.character(results[,"SVM_LABEL"])) recall_accuracy(sentiment_all[2192:3131], results[,"SVM_LABEL"]) create_precisionRecallSummary(container, results)</pre>	<ol style="list-style-type: none"> 7. Menjalankan <i>packages RTextTools</i> dan <i>e1071</i>. 8. Membuat objek kelas <i>DocumentTermMatrix</i>. 9. Mengubah data ke dalam bentuk <i>matrix</i>. 10. Membuat wadah untuk proses <i>training</i> dan <i>testing</i> data, 11. Melakukan <i>training</i> untuk mendapatkan model dengan algoritma <i>SVM</i>. 12. Menggunakan model data <i>training</i> untuk mengklasifikasin data baru. 13. Membuat tabel <i>confusion matrix</i>. 14. Menghitung nilai akurasi. 15. Menghitung nilai <i>precision</i> dan <i>recall</i>

Proses klasifikasi dilakukan dengan cara membuat *machine learning* menggunakan data latih dan data uji secara acak. Penelitian ini menggunakan metode *confusion matrix* dalam proses evaluasi. Berdasarkan pembagian data latih dan data uji yang berbeda, didapatkan hasil *confusion matrix* dengan nilai akurasi, presisi, dan *recall* yang berbeda pula yang dapat dilihat pada tabel berikut:

1. Pembagian Data *Training* 50% dan Data *Testing* 50%

Tabel 4. 17 Hasil 1

Prediksi	Aktual	
	Negatif	Positif
Negatif	597	39
Positif	115	814
Akurasi	90.15%	
<i>Precision Average</i>	89.5%	
<i>Recall Average</i>	91%	
<i>F Score Average</i>	90%	

Berdasarkan *confusion matrix* pada tabel 4.18, diketahui dari 712 ulasan negatif yang diuji, terdapat 597 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 115 ulasan yang masuk ke dalam ulasan positif. Sedangkan pada ulasan positif yang diuji, dari total 853 ulasan terdapat 814 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 39 ulasan yang masuk kedalam ulasan negatif. Kemudian dari tingkat akurasi diperoleh sebesar 90.15%, artinya dari 1565 data ulasan yang diujikan, terdapat 1411 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM).

2. Pembagian Data *Training* 60% dan Data *Testing* 40%

Tabel 4. 18 Hasil 2

Prediksi	Aktual	
	Negatif	Positif
Negatif	488	21
Positif	100	644
Akurasi	90.34%	
<i>Precision Average</i>	90%	
<i>Recall Average</i>	91,5%	
<i>F Score Average</i>	90.5%	

Berdasarkan *confusion matrix* pada tabel 4.19, diketahui dari 598 ulasan negatif yang diuji, terdapat 488 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 100 ulasan yang masuk ke dalam ulasan positif. Sedangkan pada ulasan positif yang diuji, dari total 665 ulasan terdapat 644 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 21 ulasan yang masuk ke dalam ulasan negatif. Kemudian dari tingkat akurasi diperoleh sebesar 90.34%, artinya dari 1253 data ulasan yang diujikan, terdapat 1132 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM).

3. Pembagian Data *Training* 70% dan Data *Testing* 30%

Tabel 4. 19 Hasil 3

Prediksi	Aktual	
	Negatif	Positif
Negatif	366	16
Positif	89	469
Akurasi	88.82%	
<i>Precision Average</i>	88.5%	
<i>Recall Average</i>	90%	
<i>F Score Average</i>	88.5%	

Berdasarkan *confusion matrix* pada tabel 4.20, diketahui dari 455 ulasan negatif yang diuji, terdapat 366 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 89 ulasan yang masuk ke dalam ulasan positif. Sedangkan pada ulasan positif yang diuji, dari total 485 ulasan terdapat 469 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 16 ulasan yang masuk ke dalam ulasan negatif. Kemudian dari tingkat akurasi diperoleh sebesar 88.82%, artinya dari 940 data ulasan yang diujikan, terdapat 835 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM).

4. Pembagian Data *Training* 80% dan Data *Testing* 20%

Tabel 4. 20 Hasil 4

Prediksi	Aktual	
	Negatif	Positif
Negatif	242	12
Positif	44	328
Akurasi	91.05%	
<i>Precision Average</i>	90.05%	
<i>Recall Average</i>	91.5%	
<i>F Score Average</i>	91%	

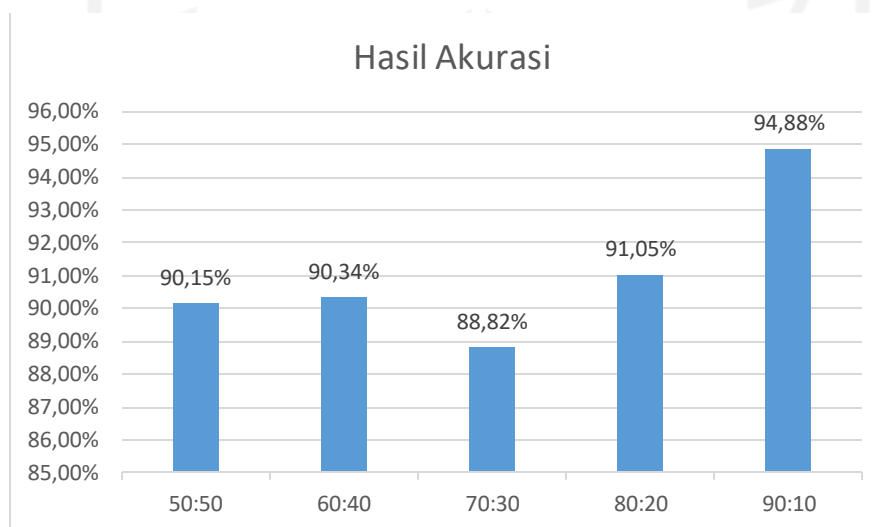
Berdasarkan *confusion matrix* pada tabel 4.21, diketahui dari 266 ulasan negatif yang diuji, terdapat 242 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 44 ulasan yang masuk ke dalam ulasan positif. Sedangkan pada ulasan positif yang diuji, dari total 330 ulasan terdapat 328 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 12 ulasan yang masuk kedalam ulasan negatif. Kemudian dari tingkat akurasi diperoleh sebesar 91.05%, artinya dari 626 data ulasan yang diujikan, terdapat 570 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM).

5. Pembagian Data *Training* 90% dan Data *Testing* 10%

Tabel 4. 21 Hasil 5

Prediksi	Aktual	
	Negatif	Positif
Negatif	124	3
Positif	13	173
Akurasi	94.88%	
<i>Precision Average</i>	94.5%	
<i>Recall Average</i>	95.5%	
<i>F Score Average</i>	94.5%	

Berdasarkan *confusion matrix* pada tabel 4.22, diketahui dari 137 ulasan negatif yang diuji, terdapat 124 ulasan yang sudah terklasifikasi dengan benar sebagai ulasan negatif dan terdapat kesalahan prediksi sebanyak 13 ulasan yang masuk ke dalam ulasan positif. Sedangkan pada ulasan positif yang diuji, dari total 176 ulasan terdapat 173 ulasan yang sudah terklasifikasi dengan benar dan terdapat kesalahan prediksi sebesar 3 ulasan yang masuk ke dalam ulasan negatif. Kemudian dari tingkat akurasi diperoleh sebesar 94.88%, artinya dari 313 data ulasan yang diujikan, terdapat 297 ulasan yang benar pengklasifikasiannya oleh model *Support Vector Machine* (SVM).



Gambar 4. 4 Grafik Perbandingan Hasil Akurasi

Berdasarkan grafik hasil akurasi diketahui bahwa hasil maksimal yang didapatkan pada perbandingan data *training* 90% dan data *testing* 10%, dengan akurasi tertinggi yaitu 94.88%.

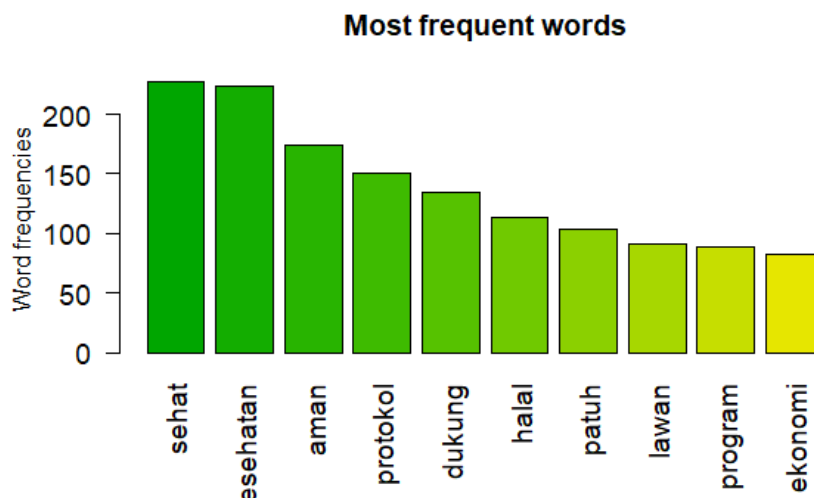
4.2.5 Visualisasi dan Asosiasi

Visualisasi dilakukan terhadap masing-masing klasifikasi kelas sentimen. Tujuan dilakukan visualisasi adalah memberikan gambaran secara umum informasi, topik, dan bahasan yang sering dibicarakan mengenai vaksin Covid-19, sehingga dari sekian banyak teks ulasan yang ada, dapat diambil informasi yang dianggap penting serta dicari asosiasi antar kata yang paling sering muncul secara bersamaan,

sehingga mampu memperkuat pencarian informasi tersebut. Berikut penjelasan hasil visualisasi dan asosiasi kata dari setiap klasifikasi kelas sentimen:

1. Ulasan Positif

Data ulasan positif yang digunakan adalah hasil pelabelan yang dilakukan baik menggunakan kamus *lexicon* maupun secara manual. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif mengenai vaksin Covid-19 yang paling sering diulas/dibicarakan. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan. Berikut merupakan visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan positif:



Gambar 4.5 Barplot Visualisasi Sentimen Positif

Berdasarkan gambar 4.5 pada *barplot* sentimen positif dengan jumlah ulasan sebanyak 1859 ulasan, diperoleh beberapa kata yang paling banyak muncul di antaranya adalah kata “sehat” dengan frekuensi sebanyak 227 kali, “kesehatan” sebanyak 223 kali, “aman” sebanyak 144 kali, “protokol” sebanyak 138 kali, “dukung” sebanyak 135 kali, “halal” sebanyak 114 kali, “patuh” sebanyak 97 kali, “lawan” sebanyak 91 kali “program” sebanyak 89 kali, dan “ekonomi” sebanyak 83 kali. Kemudian kata-kata yang muncul dibuat *wordcloud* atau kumpula kata dari ulasan yang sering digunakan oleh masyarakat mengenai vaksin Covid-19.



Gambar 4. 6 *Wordcloud* Visualisasi Sentimen Positif

Pada gambar 4.6 merupakan gambaran yang lebih jelas tentang topik dan kata-kata positif yang sering digunakan untuk membahas topik vaksin Covid-19. Semakin besar ukuran kata pada *wordcloud* menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering menggunakan kata tersebut sebagai topik pembicaraan atau penilaian positif dalam ulasan. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut:

Tabel 4. 22 Asosiasi Kata Positif

Sehat	Kesehatan	Aman
Beraktifitas (0.26)	Protokol (0.78)	Lingkungan (0.21)
Bermain (0.26)	Patuh (0.48)	Manfaatnya (0.17)
Sekolah (0.21)	Tenaga (0.25)	Protokol (0.14)
Protokol	Dukung	Halal
Kesehatan (0.78)	Mengatasi (0.44)	Mui (0.37)
Patuh (0.63)	Kepulihan (0.20)	Dibeli (0.27)
Ikuti (0.23)	Bangkitnya (0.19)	Terbukti (0.16)
Patuh	Lawan	Program
Protokol (0.63)	Adaptasi (0.64)	Dukung (0.38)
Kesehatan (0.5)	Kebiasaan (0.63)	Kepulihan (0.25)

Aman (0.26)	Sukseskan (0.19)
Ekonomi	
Pemulihan (0.49)	
Kepulihan (0.26)	
Sosial (0.23)	

Berdasarkan tabel 4.23 diperoleh asosiasi kata pada klasifikasi kelas positif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas.

Kata “sehat” berasosiasi dengan beberapa kata seperti “beraktifitas”, “bermain”, dan “sekolah”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat mengharapkan keadaan menjadi sehat kembali agar dapat melakukan aktivitas seperti bermain dan sekolah.

Kata “kesehatan” berasosiasi dengan beberapa kata seperti “protokol”, “patuh”, “tenaga”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat harus mematuhi protokol kesehatan dan tenaga kesehatan.

Kata “aman” berasosiasi dengan beberapa kata seperti “lingkungan”, “manfaatnya”, dan “protokol”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat menganggap vaksin Covid-19 aman, ramah lingkungan, dan terdapat manfaatnya. Selain itu masyarakat ingin vaksinasi Covid-19 berjalan aman dan sesuai dengan protokol.

Kata “protokol” berasosiasi dengan beberapa kata seperti “kesehatan”, “patuh”, “ikuti”. Berdasarkan asosiasi tersebut, diharapkan masyarakat mematuhi dan mengikuti protokol yang ada.

Kata “dukung” berasosiasi dengan beberapa kata seperti “mengatasi”, “kepuhian”, “bangkitnya”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat mendukung pemerintah dalam melakukan penyebaran vaksin untuk mengatasi Covid-19 dalam rangka memulihkan dan membangkitkan keadaan menjadi normal.

Kata “halal” berasosiasi dengan beberapa kata seperti “MUI”, “dibeli”, “terbukti”. Berdasarkan asosiasi tersebut, dapat diketahui vaksin Covid-19 sudah terbukti halal oleh MUI dan layak dibeli.

Kata “patuh” berasosiasi dengan beberapa kata seperti “protokol”, “kesehatan”, “aman”. Berdasarkan asosiasi tersebut, diharapkan masyarakat mematuhi protokol kesehatan agar tidak terjadi hal yang tidak diinginkan.

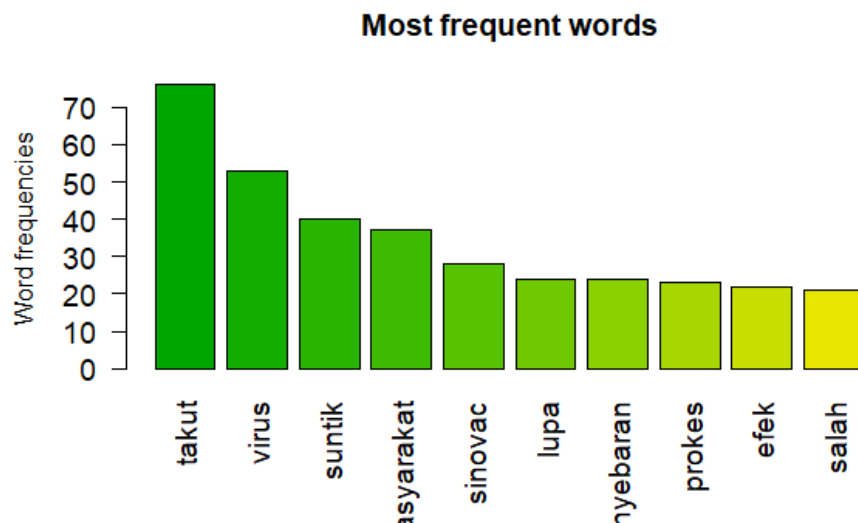
Kata “lawan” berasosiasi dengan beberapa kata seperti “adaptasi”, “kebiasaan”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat memberi saran bahwa adaptasi terlebih dahulu kebiasaan Covid, setelah itu baru memberikan perlawanan.

Kata “program” berasosiasi dengan beberapa kata seperti “dukung”, “kepulihan”, “sukseskan”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat mendukung program vaksinasi Covid-19 untuk memulihkan keadaan menjadi normal.

Kata “ekonomi” berasosiasi dengan beberapa kata seperti “pemulihan”, “kepulihan”, “sosial”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat mendukung vaksin untuk memulihkan ekonomi.

2. Ulasan Negatif

Data ulasan negatif yang digunakan adalah hasil pelabelan yang dilakukan baik menggunakan kamus *lexicon* maupun secara manual. Ekstraksi informasi pada ulasan negatif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif mengenai vaksin Covid-19 yang paling sering diulas/dibicarakan. Ulasan negatif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan. Berikut merupakan visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan negatif:



Gambar 4.7 Barplot Visualisasi Sentimen Negatif

Berdasarkan gambar 4.7 pada *barplot* sentimen negatif dengan jumlah ulasan sebanyak 1272 ulasan, diperoleh beberapa kata yang paling banyak muncul di antaranya adalah kata “takut” dengan frekuensi sebanyak 76 kali, “virus” sebanyak 56 kali, “suntik” sebanyak 40 kali, “masyarakat” sebanyak 38 kali, “sinovac” sebanyak 26 kali, “lupa” sebanyak 22 kali, “penyebaran” sebanyak 21 kali “prokes” sebanyak 20 kali, “efek” sebanyak 18 kali, dan “salah” sebanyak 17 kali. Kemudian kata-kata yang muncul dibuat *wordcloud* atau kumpulan kata dari ulasan yang sering digunakan oleh masyarakat mengenai vaksin Covid-19.



Gambar 4. 8 *Wordcloud* Visualisasi Sentimen Negatif

Pada gambar 4.8 merupakan gambaran yang lebih jelas tentang topik dan kata-kata negatif yang sering digunakan untuk membahas topik vaksin Covid-19. Semakin besar ukuran kata pada *wordcloud* menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering menggunakan kata tersebut sebagai topik pembicaraan atau penilaian negatif dalam ulasan. Selanjutnya, dilakukan pencarian asosiasi antar kata yang berhubungan dengan kata yang sering muncul secara bersamaan dan didapatkan hasil sebagai berikut:

Tabel 4. 23 Asosiasi Kata Negatif

Takut	Virus	Suntik
Jarum (0.25)	Memicu (0.36)	Gemetaran (0.32)
Suntik (0.2)	Tubuh (0.20)	Kelinci (0.24)
Efek (0.15)	Tertular (0.16)	Percobaan (0.21)
Masyarakat	Sinovac	Lupa
Kekhawatiran (0.44)	Ketidakmanjuran (0.35)	Physical (0.20)
Nolak (0.22)	Cina (0.31)	Distancing (0.20)
Bisnis (0.21)	Bisnis (0.25)	Prokes (0.26)
Penyebaran	Prokes	Salah
Waspada (0.29)	Lupa (0.21)	Dinilai (0.36)
Virus (0.21)	Disiplinkan (0.21)	Mengatasi (0.25)
Beraktivitas (0.20)	Peraturan (0.21)	Menangani (0.21)
	Efek	
	Samping (0.73)	
	Menyebabkan (0.20)	
	Diare (0.20)	

Berdasarkan tabel 4.24 diperoleh asosiasi kata pada klasifikasi kelas negatif. Proses ekstraksi informasi dengan asosiasi dilakukan secara berulang-ulang dengan cara menyaring kata-kata yang memiliki hubungan dengan kata lain dan didasarkan pada relevansi kata dengan topik yang diulas.

Kata “takut” berasosiasi dengan beberapa kata seperti “jarum”, “suntik”, dan “efek”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat belum siap untuk melakukan vaksinasi Covid-19. Hal ini dikarenakan masyarakat menganggap bahwa vaksin Covid-19 mempunyai efek samping.

Kata “virus” berasosiasi dengan beberapa kata seperti “memicu”, “tubuh”, “tertular”. Berdasarkan asosiasi tersebut, dapat diketahui protokol kesehatan yang sudah mulai dilupakan memicu tertularnya virus Covid-19.

Kata “suntik” berasosiasi dengan beberapa kata seperti “gemeteran”, “kelinci”, “percobaan”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat menganggap vaksinasi Covid-19 sebagai kelinci percobaan.

Kata “masyarakat” berasosiasi dengan beberapa kata seperti “kekhawatiran”, “nolak”, “bisnis”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat khawatir dan menolak vaksin Covid-19. Hal ini dikarenakan vaksin Covid-19 dianggap sebagai sebuah bisnis.

Kata “sinovac” berasosiasi dengan beberapa kata seperti “ketidakmanjuran”, “cina”, “bisnis”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat menganggap vaksin sinovac buatan dari cina merupakan sebuah bisnis dan tidak manjur untuk tubuh manusia.

Kata “lupa” berasosiasi dengan beberapa kata seperti “physical”, “distancing”, “prokes”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat sudah mulai lupa dengan protokol kesehatan seperti physical distancing.

Kata “penyebaran” berasosiasi dengan beberapa kata seperti “waspada”, “virus”, “beraktivitas”. Berdasarkan asosiasi tersebut, dapat diketahui penyebaran virus Covid-19 semakin banyak dan menghambat beraktivitas.

Kata “prokes” berasosiasi dengan beberapa kata seperti “lupa”, “disiplinkan”, “peraturan”. Berdasarkan asosiasi tersebut, dapat diketahui protokol kesehatan sudah mulai dilupakan dan perlu di disiplinkan lagi peraturan yang sudah berlaku.

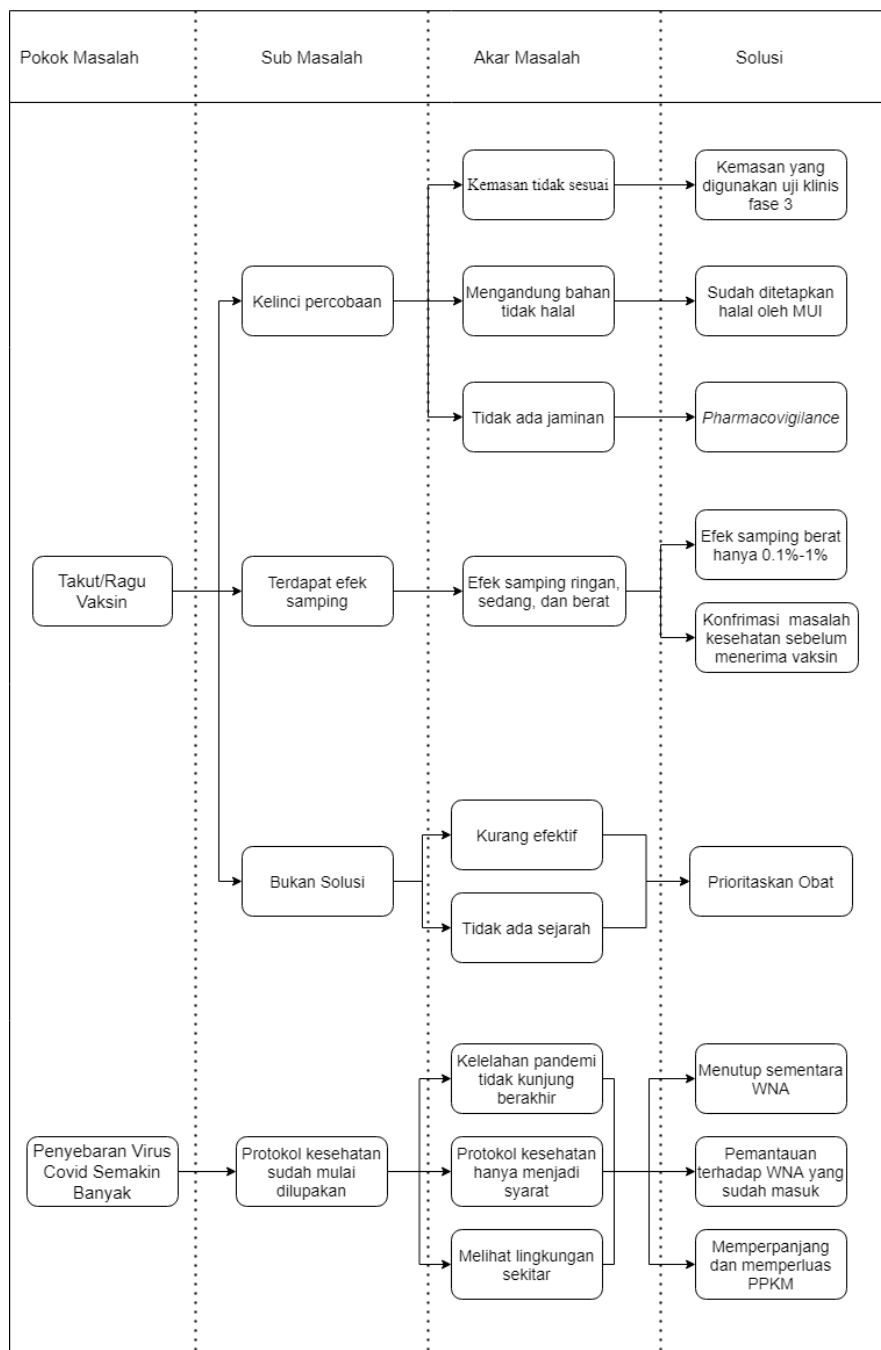
Kata “salah” berasosiasi dengan beberapa kata seperti “dimilai”, “mengatasi”, “menangani”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat menilai vaksin bukan solusi untuk menangani dan mengatasi pandemi Covid-19.

Kata “efek” berasosiasi dengan beberapa kata seperti “samping”, “menyebabkan”, “diare”. Berdasarkan asosiasi tersebut, dapat diketahui masyarakat belum siap untuk vaksinasi Covid-19. Hal ini dikarenakan vaksin tersebut menyebabkan efek samping salah satunya diare.

4.2.6 Interpretasi Hasil

Berdasarkan masalah yang didapatkan dari hasil ulasan negatif, kemudian dicari akar dari masalah tersebut dan didapatkan solusi yang ada pada sebuah literatur. Setelah itu dianalisis dengan diagram pohon sebagai berikut:





Gambar 4. 9 Diagram Pohon

BAB V

PEMBAHASAN

5.1 Gambaran Umum Persepsi masyarakat

Berdasarkan data ulasan masyarakat pada bulan Januari dan bulan februari yang terlihat pada gambar 4.2 didapatkan jumlah data sebanyak 3131 ulasan. Pembagian klasifikasi data terhadap ulasan masyarakat dibagi menjadi tiga kelas, yaitu sentimen positif sebanyak 1735 data ulasan, sentimen netral sebanyak 160 data ulasan, dan sentimen negatif sebanyak 1236 data ulasan. Dikarenakan yang digunakan hanya sentimen positif dan negatif, maka dari itu kelas netral direduksi lagi untuk mengetahui kalimat tersebut termasuk sentimen positif atau negatif. Hal ini dikarenakan sentimen netral kurang memberikan masukan dan manfaat bagi pihak pemerintah. Dalam mereduksi kelas sentimen netral yang mengandung unsur positif seperti sehat, halal dimasukkan ke dalam kategori sentimen positif, sedangkan yang mengandung unsur negatif seperti takut, ragu dimasukkan ke dalam kategori sentimen negatif.

Data ulasan sentimen netral setelah direduksi maka jumlah sentimen positif menjadi 1859 ulasan, sedangkan untuk kelas sentimen negatif menjadi 1272 ulasan. kemudian seperti terlihat pada gambar 4.3 dapat diketahui bahwa pada bulan Januari tanggapan negatif lebih besar jika dibandingkan dengan tanggapan positif. Kemudian untuk bulan Februari tanggapan positif lebih besar jika dibandingkan dengan tanggapan negatif. Hal ini terjadi mungkin dikarenakan pada bulan Januari vaksin Covid-19 baru disebarkan kepada masyarakat yang menyebabkan masyarakat kurang percaya dengan vaksin tersebut. Akan tetapi dapat dilihat pada gambar 4.3 khusus nya pada bulan Januari jumlah sentimen negatif dan positif

perbedaannya tidak terlalu jauh. Hal ini mungkin disebabkan pada saat tanggal 13 Januari presiden Jokowi melakukan vaksin pertama kali dihadapan masyarakat republik Indonesia yang memberikan kepercayaan kepada masyarakat. Kemudian untuk bulan february masyarakat sudah mulai percaya dengan vaksinasi Covid-19, hal ini terbukti dengan adanya vaksinasi di setiap kota-kota besar seperti Yogyakarta.

5.2 Hasil Penerapan Metode *Support Vector Machine* Dalam Proses Klasifikasi Ulasan Sentimen

Klasifikasi menggunakan metode *Support Vector Machine* dilakukan dengan menggunakan 5 kombinasi data *training* dan data *testing* antara lain 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10% dari total data. Akurasi yang didapatkan dari pengujian data tes yang tertinggi pada saat pengujian kelima yaitu 94.88%, diikuti pengujian keempat dengan akurasi sebesar 91.05%, pengujian kedua dengan akurasi sebesar 90.34%, pengujian pertama dengan akurasi sebesar 90.15%, dan akurasi terendah pada saat pengujian ketiga dengan akurasi sebesar 88.82%.

Precision tertinggi pada saat pengujian kelima sebesar 94.5%, diikuti pengujian keempat sebesar 90.05%, pengujian kedua sebesar 90%, diikuti pengujian pertama dengan sebesar 89.5%, dan *precision* terendah pada saat pengujian ketiga sebesar 88.5%.

Recall tertinggi pada saat pengujian kelima sebesar 95.5%, diikuti pengujian keempat dan kedua sebesar 91.05%, diikuti pengujian pertama dengan sebesar 91%, dan *recall* terendah pada saat pengujian ketiga sebesar 90%.

F score tertinggi pada saat pengujian kelima sebesar 94.5%, diikuti pengujian keempat sebesar 91%, pengujian kedua sebesar 90%, diikuti pengujian pertama dengan sebesar 90.5%, dan *f score* terendah pada saat pengujian ketiga sebesar 88.5%.

Kesalahan dalam proses klasifikasi dapat terjadi jika disebabkan oleh penggunaan dataset yang kurang tepat. Keadaan ini yang menyebabkan banyak ditemukan fitur kemunculan data yang bukan termasuk kategorinya pada data uji yang digunakan. Contohnya pada dataset yang digunakan dalam data positif terdapat kata “ragu”, kemudian pada data negatif terdapat kata “aman”, dimana keduanya saling bertolak belakang.

5.3 Hasil Visualisasi dan Asosiasi Kata

Berdasarkan hasil ulasan positif dari hasil pelabelan, terdapat 1859 ulasan positif mengenai vaksinasi Covid-19. Kemudian dari gambar 4.5 terdapat beberapa kata yang sering muncul pada ulasan positif tersebut. Beberapa kata yang paling sering muncul diantaranya adalah kata “sehat” dengan frekuensi sebanyak 227 kali, “kesehatan” sebanyak 223 kali, “aman” sebanyak 144 kali, “protokol” sebanyak 138 kali, “dukung” sebanyak 135 kali, “halal” sebanyak 114 kali, “patuh” sebanyak 97 kali, “lawan” sebanyak 91 kali “program” sebanyak 89 kali, dan “ekonomi” sebanyak 83 kali.

Berdasarkan asosiasi kata pada sentimen kelas positif, dapat diketahui informasi yang didapatkan dari analisis asosiasi kata kelas positif adalah sebagai berikut. Pertama vaksin Covid-19 sudah terbukti halal oleh MUI. Kemudian yang kedua masyarakat mendukung pemerintah dalam melakukan vaksinasi Covid-19 dalam rangka mengembalikan keadaan menjadi sehat kembali. Selanjutnya yang ketiga vaksin Covid-19 sudah terbukti aman dan diharapkan tetap patuh terhadap protokol kesehatan.

Sedangkan, berdasarkan hasil ulasan negatif dari hasil pelabelan, terdapat 1272 ulasan negatif mengenai vaksinasi Covid-19. Kemudian dari gambar 4.7 terdapat beberapa kata yang sering muncul pada ulasan negatif tersebut. Beberapa kata yang sering muncul diantaranya adalah kata “takut” dengan frekuensi sebanyak 76 kali, “virus” sebanyak 56 kali, “suntik” sebanyak 40 kali, “masyarakat” sebanyak 38 kali, “sinovac” sebanyak 26 kali, “lupa” sebanyak 22 kali, “penyebaran” sebanyak 21

kali “prokes” sebanyak 20 kali, “efek” sebanyak 18 kali, dan “salah” sebanyak 17 kali.

Berdasarkan asosiasi kata pada sentimen kelas negatif, dapat diketahui informasi yang didapatkan dari analisis asosiasi kata kelas negatif adalah sebagai berikut. Pertama masyarakat masih takut untuk melakukan vaksinasi Covid-19, hal ini dikarenakan masih terdapat persepsi terkait kelinci percobaan, efek samping, dan bukan solusi yang tepat, kemudian penyebaran virus Covid-19 semakin banyak, hal ini dikarenakan protokol kesehatan sudah mulai dilupakan dan perlu didisiplinkan lagi peraturan yang berlaku.

5.4 Penjelasan Diagram Pohon

Berdasarkan diagram pohon pada gambar 4.9 diketahui terdapat 2 pokok permasalahan yaitu takut/ragu vaksin, dan penyebaran virus Covid semakin banyak. Kemudian terdapat 3 sub masalah pada pokok masalah takut/ragu vaksin yaitu kelinci percobaan, terdapat efek samping, dan bukan solusi. Sedangkan pada pokok masalah penyebaran virus Covid semakin banyak hanya terdapat 1 satu sub masalah yaitu protokol kesehatan sudah mulai dilupakan. Berikut merupakan penjelasan mengenai akar permasalahan dan solusi setiap masalah yang ada berdasarkan berita yang ada:

Tabel 5. 1 Pembahasan Diagram Pohon

Akar Masalah	Solusi
Kelinci Percobaan	
<ul style="list-style-type: none"> Kemasan Tidak Sesuai <p>Pada tanggal 2 Januari 2021 kemasan vaksin Sinovac berlabel “<i>only for clinical trial</i>”.</p>	<ul style="list-style-type: none"> Kemasan Yang Digunakan Uji Klinis Fase 3 <p>Menurut Sekretaris Perusahaan Bio Farma Bambang Heriyanto menjelaskan, bahwa kemasan sinovac yang terdapat tulisan “<i>only for clinical</i>”</p>

Akar Masalah	Solusi
	<p><i>trial</i>” tersebut adalah kemasan yang digunakan untuk vaksin yang dipakai dalam uji klinis fase 3 atau yang digunakan saat penelitian, bukan vaksin yang nantinya akan disuntikkan kepada masyarakat.</p>
<ul style="list-style-type: none"> • Mengandung Bahan Tidak Halal Pada kemasan tersebut dituliskan <i>composition</i> dan <i>description</i> berasal dari Vero Cell atau berasal dari jaringan Kera hijau Afrika. 	<ul style="list-style-type: none"> • Sudah Ditetapkan Halal Oleh MUI Pada tanggal 11 Januari 2021 vaksin Sinovac hukumnya suci dan halal dan menurut Sekretaris Perusahaan Bio Farma Bambang Heriyanto menjelaskan, Vero Cell yang disebut pada informasi yang beredar adalah media tumubuh kebang dari virus corona. Virus perlu ditumbuhkan dalam Vero Cell karena virus hanya tumbuh di sel hidup.
<ul style="list-style-type: none"> • Tidak Ada Jaminan Masyarakat khawatir jika setelah vaksinasi terkena efek samping dari vaksin Covid-19. 	<ul style="list-style-type: none"> • <i>Pharmacovigilance</i> Yang dimaksud <i>pharmacovigilance</i> adalah seluruh kegiatan tentang pendeteksian, penilaian, pemahaman, dan pencegahan efek samping atau masalah lainnya terkait penggunaan obat yang dimonitoring oleh lembaga independen Komisi Nasional Kejadian Ikutan Paska Imunisasi (KIPI).
Terdapat Efek Samping	
<ul style="list-style-type: none"> • Efek Samping Ringan, Sedang, Berat 	<ul style="list-style-type: none"> • Efek Samping Berat Hanya 0.1%-1%

Akar Masalah	Solusi
<p>Masyarakat khawatir terhadap efek samping vaksin Sinovac yang belum ditemukan.</p>	<p>Menurut Kepala BPOM Penny Lukito Efek samping derajat berat yang dilaporkan hanya 0,1% hingga 1%. Efek samping itu tidak berbahaya dan bisa pulih kembali dan menurut anggota Komisi IX DPR dari Fraksi Partai Golkar Yahya Zaini keamanan vaksin Sinovac sudah dijamin. Bahkan efikasinya mencapai 63%, di atas standar minimal yang dipersyaratkan WHO sebesar 50%.</p>
Bukan Solusi	
<ul style="list-style-type: none"> • Kurang Efektif <p>Hal ini dikarenakan menggunakan vaksin masih dapat tertular oleh virus Covid-19. Kemudian mutasi virus Covid-19 membuat efektifitas vaksin tidak terjamin.</p>	<ul style="list-style-type: none"> • Prioritaskan Obat <p>Menurut Mantan Menteri Kesehatan Siti Fadilah Supari menjelaskan, bahwa pandemi terjadi karena penyakit yang menyebar belum di temukan obatnya. Maka dari itu obat lebih penting daripada vaksin, karena vaksin seharusnya dibuat saat virus yang menyebar telah stabil.</p>
<ul style="list-style-type: none"> • Tidak Ada Sejarahnya <p>Menurut Mantan Menteri Kesehatan Siti Fadilah Supari menjelaskan, dalam sejarah tidak yang mengatakan pandemi dapat dihentikan oleh vaksin.</p>	
Protokol Kesehatan Sudah Mulai Dilupakan	
<ul style="list-style-type: none"> • Kelelahan Pandemi Tidak Kunjung Berakhir 	<ul style="list-style-type: none"> • Menutup Sementara WNA

Akar Masalah	Solusi
<p>Menurut Ketua Subbidang Sosialisasi Perubahan Perilaku Satgas COVID-19/Deputi Dalduk BKKBN, Dr. Ir. Dwi Listyawardani menjelaskan, meski masyarakat hanya diminta untuk menerapkan 3M, namun pandemi yang tak kunjung berakhir, membuat akhirnya lelah mematuhi protokol kesehatan</p>	<p>Ketua Komisi A DPRD DKI Jakarta Mujiyono mengatakan terdapat beberapa aspek yang perlu dilakukan penajaman, salah satunya adalah melanjutkan kebijakan untuk menutup warga negara asing (WNA).</p> <ul style="list-style-type: none"> • Pemantauan Terhadap WNA Yang Sudah Masuk <p>Kemudian yang kedua yaitu pemantauan terhadap WNA yang sudah terlanjur masuk ke Indonesia sebelum dilakukan penutupan.</p>
<ul style="list-style-type: none"> • Protokol Kesehatan Hanya Menjadi Syarat <p>Menurut Sosiolog dari Universitas Sebelas Maret (UNS) Dr. Drajat Tri Kartono, MSi, menjelaskan bahwa, protokol kesehatan kini hanya dijadikan sebagai syarat atau salah satu norma untuk mengadakan berbagai aktivitas, ritual budaya, perjalanan, pekerjaan, dan aktivitas pendidikan.</p>	<ul style="list-style-type: none"> • Memperpanjang dan Memperluas PPKM <p>PPKM telah dilaksanakan mulai tanggal 11 Januari 2021 dan diperpanjang hingga 31 Mei 2021. Kemudian yang sebelumnya hanya 7 provinsi, sekarang menjadi 30 provinsi.</p>
<ul style="list-style-type: none"> • Melihat Lingkungan Sekitar <p>Menurut Menteri Dalam Negeri (Mendagri) Tito Karnavian menjelaskan bahwa lingkungan juga berpengaruh terhadap disiplinnya protokol kesehatan.</p>	

BAB VI

PENUTUP

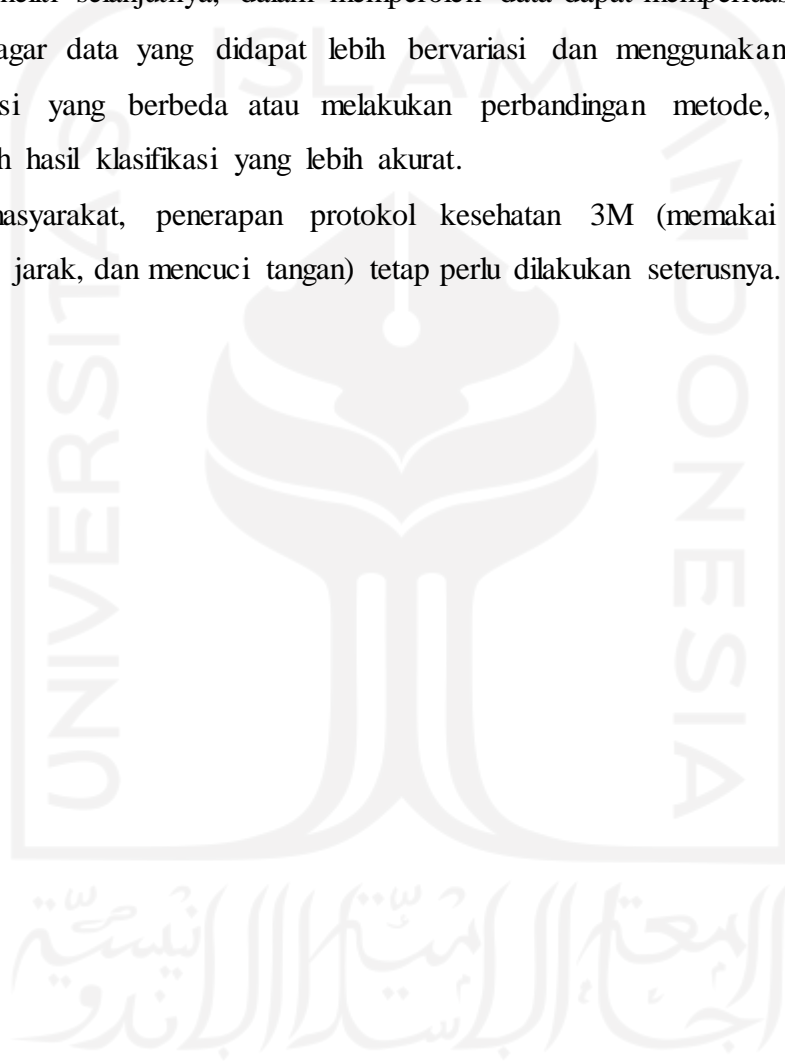
6.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, kesimpulan dalam penelitian ini adalah sebagai berikut:

1. Berdasarkan persepsi terkait dengan vaksin Covid-19 pada periode Januari-Februari 2021 terdapat 3131 data. Kemudian pelabelan kelas sentimen didapatkan jumlah kelas positif adalah sebanyak 1859 ulasan, yang terdiri dari bulan Januari sebanyak 837 ulasan dan bulan Februari sebanyak 1022 ulasan dan kelas negatif sebanyak 1272 ulasan, yang terdiri dari bulan Januari sebanyak 869 ulasan dan bulan Februari 403 ulasan
2. Berdasarkan perbandingan data *training* dan data *testing* yang dilakukan sebanyak 5 kali percobaan dengan perbandingan yang berbeda, didapatkan akurasi terbaik pada perbandingan 90% : 10% dengan akurasi 94.88%.
3. Berdasarkan asosiasi dan visualisasi negatif, dapat diketahui terdapat dua pokok masalah yaitu takut/ragu vaksin dan penyebaran virus Covid-19 semakin banyak.

6.2 Saran

1. Bagi pemerintah, menjadi bahan evaluasi terhadap pihak yang terkait dalam melakukan pendistribusian vaksin Covid-19.
2. Bagi peneliti selanjutnya, dalam memperoleh data dapat memperluas periode waktu agar data yang didapat lebih bervariasi dan menggunakan metode klasifikasi yang berbeda atau melakukan perbandingan metode, sehingga diperoleh hasil klasifikasi yang lebih akurat.
3. Bagi masyarakat, penerapan protokol kesehatan 3M (memakai masker, menjaga jarak, dan mencuci tangan) tetap perlu dilakukan seterusnya.



Daftar Pustaka

- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia*.
- Abdurahman, H., & Riswaya, A. . (2014). Aplikasi Pembayaran Secara Kredit Pada Bank Yudha Bahakti. *Jurnal Computech & Bisnis*, 8(2), 61–69.
- Aliandu, P. (2012). *Analisis Sentimen Tweet Berbahasa Indonesia di Twitter*. Yogyakarta: Universitas Gadjah Mada.
- Anggreini, D. (2008). *Klasifikasi Topik Menggunakan Metode Naive Bayes dan Maximum Entropy pada Artikel Media Massa dan Abstrak Tulisan*. Universitas Indonesia.
- Aprilla, S., Furqon, M. T., & Fauzi, M. A. (2018). Klasifikasi Penyakit Skizofrenia dan Episode Depresi Pada Gangguan Kejiwaan Dengan Menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*.
- Azizah, K. N. (2021, Januari 25). *Indonesia Mulai Vaksinasi COVID-19, Menkes: Malaysia Ngomel Belum Dapat*. Retrieved Januari 27, 2021, from detikhealth: <https://health.detik.com/berita-detikhealth/d-5347089/indonesia-mulai-vaksinasi-covid-19-menkes-malaysia-ngomel-belum-dapat>
- Bramer, M. (2007). *Principles of Data Mining: Undergraduate Topics in Computer Science*. London: Springer-Verlag.
- Buntoro, G. A. (2016). Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *Jurnal Dinamika Informatika*, 5(2).
- Choy, M. (2019). A Sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census corection. *arXiv preprint arXiv: 1108.5520*.
- Fadlisyah. (2014). *Statistika : Terapannya di Informatika, 1st ed.* Yogyakarta: Graha Ilmu.
- Falahah, & Nur, D.D. (2015). Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naive Bayes. *Seminar Nasional Sistem Informasi Indonesia*, 336-340.
- Fanani, F. (2017). *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*.
- Faradhillah, N. Y., Kusumawardani, R. P., & Hafidz, I. (2016). Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin. *SESINDO 2016*.
- Fathan Hidayatullah, A., & Sn, A. (2014). ISSN: 1979-2328 UPN "Veteran. *Seminar Nasional Informatika, 2014(semnasIF)*, 115–122.

<http://www.situs.com>

- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 861–874.
- Feldman, R., dan Sanger, J. (2006). *The Text Mining Handbook*.
- Fritz, G. (2016). *Analisa Bad Hike Pada Kran Lavatory Tipe S11234R Menggunakan Metode Nominal Group Technique dan Metode Fishbone di PT Surya Toto Indonesia tbk*. Universitas Gadjah Mada.
- Gumilang, Z. (2018). *Implementasi Naive-Bayes Classifier dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-commerce Shopee pada Situs Google*. Yogyakarta: Program Studi Statistika FMIPA Universitas Islam Indonesia.
- Han, J., & Kamber, M. (2006). . *Data Mining: Concepts and Technuques Second Eddition*. USA: Elsevier.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and Techniques Second Edition*. San Fransisco: Morgan Kauffman.
- Josi, A., Abdillah, L., & Suryayusra. (2014). Penerapan Teknik Web Scraping Pada MesinPencari Artikel Ilmiah. *Jurnal Sistem Informasi (SISFO)*, 05.
- Kemenkes, R. (2015). *Profil Kesehatan Indonesia tahun 2014*. Jakarta: Kemenkes RI.
- Kohavi, R., & Provost, F. (1998). On applied research in machine learning. *Machine Learning*,, 127-132.
- Kothainayaki, S., dkk. (2012). User Preferences on University Websites: A Study. *Library Philosophy and Practice. (e-journal)*, 788.
- Kristiyanti, D. A. (2015). Analisis sentimen review produk kosmetik melalui komparasi feature selection. *Konferensi Nasional Ilmu Pengetahuan Dan Teknologi (KNIT)*,, 74-81.
- Maulana, A. (2017, September 27). *Cuitan di Twitter kini bisa 280 karakter*. Retrieved Desember 1, 2020, from <https://www.cnnindonesia.com>
- Napitupulu, D. (2017). Analysis of Factors Affecting Website Quality Based on Webqual Approach (Study Case: XYZ University). *Advanced Science Engineering Information Technology*, 792.
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *EProceedings of Engineering*, 2(1).
- Nugroho, A. ., Wranto, A. ., & Handoko, D. (2003). *Support Vector Machine Teori*

dan Aplikasinya dalam Bioinformatika. IlmuKomputer.com.

- O'Keefe, & Koprinska. (2009). Feature selection and weighting methods in sentiment analysis. *Proceedings of the 14th Australasian Document Computing Symposium Sydney*.
- Pakar Online Indonesia. (2012, Desember 30). *Mengenal Twitter dan Daftar Istilah yang Digunakan*. Retrieved Desember 1, 2020, from <https://www.pakaronline.com/twitter/mengenal-twitter-dan-daftar-istilah-yang-digunakan/>
- Pang, Bo, Lee, Lilian. (2008). Opinion Mining and Sentiment Analysis . *Foundation and Trend in Information Retrieval*.
- Pintoko, B. M., & Lhaksana, K. M. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier. *EProceedings of Engineering*, 5(3).
- Plotkin, S. (2013). *Vaccine Fact Book*. Pennsylvania: PhRMA.
- Prakoso, R. ., Novianti, A., & Setianingsih, C. (2017). Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy. *E-Proceeding of Engineering*, 4(2), 2389–2395.
- Prasetyo, E. (2012). *Data mining Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi.
- Pratama, A. M. (2020, September 11). *Pengguna Internet Indonesia hingga Kuartal II 2020 Capai 196,7 Juta Orang*. Retrieved from Kompas.com: <https://money.kompas.com/read/2020/11/09/213534626/pengguna-internet-indonesia-hingga-kuartal-ii-2020-capai-1967-juta-orang#:~:text=JAKARTA%2C%20KOMPAS.com%20%2D%20Jumlah,jiwa%20hingga%20kuartal%20II%202020.&text=Hal%20tersebut%20diketahui%20berdasarkan%2>
- Proverawati, A dan Andhini C.S.D. (2010). *Imunisasi dan Vaksinasi*. Yogyakarta: Nuha Offset.
- Putranti, N. ., & Winarko, E. (2014). Analisis sentimen twitter untuk teks berbahasa Indonesia dengan maximum entropy dan support vector machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91–100.
- Putri, D. (2016). *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Universitas Gadjah Mada.
- Putri, G. S. (2020, December 22). *Banyak Orang Ragu Terhadap Vaksin Covid-19, Kenapa Bisa Terjadi?* Retrieved Januari 27, 2021, from KOMPAS.COM: <https://www.kompas.com/sains/read/2020/12/22/130300423/banyak-orang-ragu-terhadap-vaksin-covid-19-kenapa-bisa-terjadi-?page=all>

- Rahayu, U. (2020). *Pro Kontra Rencana Vaksin COVID-19 di Indonesia*. Hellosehat. <https://hellosehat.com/infeksi/covid19/rencana-vaksin-covid-19/>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). *Cross Validation*. New York: Encyclopedia of Database Systems, Springer.
- Riyanto, A. D. (2020, January). *Hootsuite (We are Social): Indonesian Digital Report 2020*. Retrieved Desember 10, 2020, from [andi.link: https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2020/](https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2020/)
- Rozi, I. F. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrical Power, Electronics, Communications, Controls, and Informatics Seminar (EECCIS)*, 37-43.
- Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 1(12), 1725–1732. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>
- Rustian, R. (2019, Maret 1). *Apa itu Social Media*. Retrieved Desember 1, 2020, from www.unpas.ac.id
- Shaver, P. R., Mudaya, U., & Fraley, R. C. (2001). Structure of Indonesian Emotion Lexicon. *Asian Journal of Psychology*.
- Silverman, S. N., & Lori L, S. (1994). "Using total quality tools for marketing research: A qualitative approach for collecting, organizing, and analyzing verbal response data." *Advanced Research Techniques Forum*.
- Sunarti. (2012). *Pro Kontra Imunisasi: "Manfaat Imunisasi"*. Yogyakarta: Hanggar Kreator.
- Tan, P. e. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- Tribunnews. (2021). *Deretan Tokoh hingga Publik Figur yang Tolak Vaksin Covid-19, Politikus hingga Adik Supermodel Artikel ini telah tayang di Tribunnews.com dengan judul Deretan Tokoh hingga Publik Figur yang Tolak Vaksin Covid-19, Politikus hingga Adik Supermodel*, <https://www.tribunnews.com/nasional/2021/01/20/deretan-tokoh-hingga-publik-figur-yang-tolak-vaksin-covid-19-politikus-hingga-adik-supermodel>
- Turban, E., Aronson, J. E., & Liang, T. P. (2005). *Decision Support Systems and*. Yogyakarta: Andi.
- Turland, M. (2010). *Guide to Web Scraping with PHP*.

- Ulwan, M. N. (2016). *Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine dan Association*. Universitas Islam Indonesia.
- USK, S. C.-1. (2020, Maret 11). *Pusat Data dan Informasi Universitas Syiah Kuala Satgas Covid-19*. Retrieved Januari 27, 2021, from <http://covid19.unsyiah.ac.id/>: [http://covid19.unsyiah.ac.id/latar-belakang/#:~:text=Penyakit%20corona%20virus%202019%20atau,2%20\(SARS%2DCOV2\).](http://covid19.unsyiah.ac.id/latar-belakang/#:~:text=Penyakit%20corona%20virus%202019%20atau,2%20(SARS%2DCOV2).)



Lampiran

Lampiran 1. Scraping Data

```
#Download sertifikat dari curl
download.file(url="http://curl.haxx.se/ca/cacert.pem", destfile = "cacert.pem")

#Meminta izin kepada twitter dengan meruning berikut:
reqURL <-"https://api.twitter.com/oauth/request_token"
accessURL <-"https://api.twitter.com/oauth/access_token"
CUSTOMER_KEY <- "TssLr2b4ySsTPq02WKweVYQBN"
CUSTOMER_SECRET <-"ompeZRYM7y4Vk8A33T2m3nAPIgzoaCljYvuEEpdz0LwsvI8ZY4"
ACCESS_TOKEN <- "410688726-JOI4A4vjcwC0SkbsdxmQCeIVGAdpHjY7v3HiLA9V"
ACCESS_Secret <-"82ZXTGOEJHfIfYaaEbgM3Jlwf2MuiIfPsDITRldEOPk44"

#Men setup authorization
setup_twitter_oauth(CUSTOMER_KEY, CUSTOMER_SECRET, ACCESS_TOKEN,
ACCESS_Secret)
1

#Mengambil tweet tentang vaksin dengan filter language versi bahasa indonesia
search.string<-"vaksin"
no.of.tweets <- 100

#Versi bahasa indonesia
vaksin.Tweets <- searchTwitter(search.string, n=no.of.tweets,lang="id")
df_id <- do.call("rbind", lapply(vaksin.Tweets, as.data.frame))
view(df_id)
write.csv(df_id,"C:\\Users\\hp\\Documents\\andika4.csv",row.names = TRUE)
```

Lampiran 2. Preprocessing

```
# INSTALL PACKAGES
install.packages("tm") # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
```

```

install.packages("RColorBrewer") # color palettes

# LOAD PACKAGES
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library(stringr)

setwd("E:\\KULIAH\\METOPEL\\SKRIPSI\\")
docs <- readLines("vaksin.csv")

# LOAD THE DATA AS A CORPUS
docs <- Corpus(VectorSource(docs))
inspect(docs)

# REPLACING “/”, “@” AND “|” WITH SPACE
toSpace <- content_transformer(function (x, pattern) gsub (pattern, " ",x))
docs <- tm_map(docs,toSpace, "/")
docs<-tm_map(docs,toSpace, "!")
docs<-tm_map(docs,toSpace, "@")

# CLEANING THE TEXT TO LOWER CASE
docs <- tm_map(docs, content_transformer(tolower))
# REMOVE PUNCTUATION
docs <- tm_map(docs, toSpace, "[[:punct:]]")

# REMOVE NUMBER
docs <- tm_map(docs, toSpace, "[[:digit:]]")

# REPLACE WORDS
docs <- tm_map(docs, gsub, pattern="jdi", replacement = "jadi")
docs <- tm_map(docs, gsub, pattern="tdk", replacement = "tidak")
docs <- tm_map(docs, gsub, pattern="yg", replacement = "yang")
docs <- tm_map(docs, gsub, pattern="mantul", replacement = "mantap")
docs <- tm_map(docs, gsub, pattern="mntap", replacement = "mantap")
docs <- tm_map(docs, gsub, pattern="gmn", replacement = "bagaimana")

```

```

docs <- tm_map(docs, gsub, pattern="cpt", replacement = "cepat")
docs <- tm_map(docs, gsub, pattern="bgt", replacement = "banget")
docs <- tm_map(docs, gsub, pattern="dtg", replacement = "datang")
docs <- tm_map(docs, gsub, pattern="problem", replacement = "masalah")
docs <- tm_map(docs, gsub, pattern="tlg", replacement = "tolong")
docs <- tm_map(docs, gsub, pattern="jd", replacement = "jadi")
docs <- tm_map(docs, gsub, pattern="sdh", replacement = "sudah")
docs <- tm_map(docs, stem_text)
inspect(docs[1:10])

stopwordindo = readLines("stopwords.csv")

# REMOVE STOPWORDS FROM CORPUS
docs <- tm_map(docs, removeWords, stopwordindo)

# REMOVE YOUR OWN STOPWORDS
# SPECIFY YOUR STOPWORDS AS A CHARACTER VECTOR
docs <- tm_map(docs, removeWords,
c("vaksin", "covid", "pandemi", "dan", "presiden", "corona", "kali", "saya", "gue", "aku", "ini", "divaksin",
"vaksinasi", "kok", "aja", "nya"))

# REMOVE URL
docs <- tm_map(docs, removeURL)
removeURL <- function(x) gsub("http[[:alnum:]]*", "", x)

# ELIMINATE EXTRA WHITE SPACES
docs <- tm_map(docs, stripWhitespace)
inspect(docs)
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m), decreasing = TRUE)
d <- data.frame(word=names(v), freq=v)
head(d,20)

# BUILD A TERM-DOCUMENT MATRIX
dataframe <- data.frame(text=unlist(sapply(docs, `[]`)), stringsAsFactors=F)
write.csv(dataframe, "E:/KULIAH/MET OPEL/SKRIPSI/preprocessing1.csv")

```

Lampiran 3. Pelabelan

```

#MENJALANKAN PACKAGES
library(tm)
library(plyr)

#MENGATUR DIREKTORI KERJA
setwd("E:\\KULIAH\\METOPEL\\SKRIPSI\\")
kalimat2<-read.csv("preprocessing1.csv", header=TRUE)

#Scoring data berdasarkan kamus kata
positif <- scan("positif.txt",what="character",comment.char=";")
negatif <- scan("negatif.txt",what="character",comment.char=";")
kata.positif = c(positif)
kata.negatif = c(negatif)
score.sentiment = function(kalimat2, kata.positif, kata.negatif,
                           .progress='none')
{
  require(plyr)
  require(stringr)
  scores = laply(kalimat2, function(kalimat, kata.positif, kata.negatif)
  {
    kalimat = gsub("[[:punct:]]", "", kalimat)
    kalimat = gsub("[[:cntrl:]]", "", kalimat)
    kalimat = gsub("\\d+", "", kalimat)
    kalimat = tolower(kalimat)
    list.kata = str_split(kalimat, "\\s+")
    kata2 = unlist(list.kata)
    positif.matches = match(kata2, kata.positif)
    negatif.matches = match(kata2, kata.negatif)
    positif.matches = !is.na(positif.matches)
    negatif.matches = !is.na(negatif.matches)
    score = sum(positif.matches) - (sum(negatif.matches))
    return(score)
  }, kata.positif, kata.negatif, .progress=.progress )
  scores.df = data.frame(score=scores, text=kalimat2)

```

```

return(scores.df)
}
hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)
head(hasil)
View(hasil)

#CONVERT SCORE TO SENTIMENT
hasil$klasifikasi<- ifelse(hasil$score<0, "Negatif", "Positif")

hasil$klasifikasi
View(hasil)

#Tukar Row
data <- hasil[c(3,1,2)]
View(data)

write.csv(data, file = "E:\\KULIAH\\METOPEL\\SKRIPSI\\PELABELAN2.csv")
#Menyimpan Data Positif dan Negatif
data.pos <- hasil[hasil$score>0,]
View(data.pos)
write.csv(data.pos, file = "E:\\KULIAH\\METOPEL\\SKRIPSI\\positive.csv")
data.neg <- hasil[hasil$score<0,]
View(data.neg)
write.csv(data.neg, file = "E:\\KULIAH\\METOPEL\\SKRIPSI\\negative.csv")

```

Lampiran 4. Script R Visualisasi dan Asosiasi Kata

```

# LOAD PACKAGES
library(NLP)
library(tm)
library(SnowballC)
library(RColorBrewer)
library(wordcloud)
library(stringr)
library(wordcloud2)

# INPUT POSITIVE SENTIMENT DATA
setwd("E://KULIAH/METOPEL/SKRIPSI")

```



```

docs <- read.csv("positive.csv")

# LOAD THE DATA AS A CORPUS
docs<- Corpus(VectorSource(docs))

docs <- tm_map(docs, stripWhitespace)
docs[[1]][1]

# BUILD A TERM-DOCUMENT MATRIX
dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 70)

# GENERATE THE WORDCLOUD
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=50, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

# EXPLORE FREQUENT TERMS AND THEIR ASSOCIATIONS
findFreqTerms(dtm, lowfreq = 4)

# WORD ASSOCIATION
v<-as.list(findAssocs(dtm, terms =c("sehat","kesehatan", "aman", "protokol","dukung", "halal",
"patuh",
"lawan", "program", "ekonomi"),corlimit
c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)))
v

# BARPLOT
k<-barplot(d[1:10,]$freq, las = 2, names.arg =
          d[1:10,]$word,cex.axis=1.2,cex.names=1.2,
          main ="Most frequent words",
          ylab = "Word frequencies",col = terrain.colors(20))
termFrequency<- rowSums(as.matrix(dtm))

```

```

termFrequency<- subset(termFrequency,termFrequency>=48)
text(k,sort(termFrequency, decreasing = T)-
      1,labels=sort(termFrequency, decreasing = T),pch = 6, cex = 1)

# INPUT NEGATIVE SENTIMENT DATA
setwd("E://KULIAH/METPEL/SKRIPSI")
docs <- read.csv("negative.csv")

# LOAD THE DATA AS A CORPUS
docs<- Corpus(VectorSource(docs))
docs <- tm_map(docs, stripWhitespace)
docs[[1]][1]

# BUILD A TERM-DOCUMENT MATRIX
dtm<- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 70)

# GENERATE THE WORDCLOUD
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=50, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))

# EXPLORE FREQUENT TERMS AND THEIR ASSOCIATIONS
findFreqTerms(dtm, lowfreq = 4)

# WORD ASSOCIATION
v<-as.list(findAssocs(dtm, terms =c("takut","virus", "suntik", "masyarakat","sinovac", "lupa",
"penyebaran",
"prokes", "efek", "salah"),corlimit
=c(0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)))
V

# BARPLOT

```

```

k<-barplot(d[1:10,]$freq, las = 2, names.arg =
  d[1:10,]$word,cex.axis=1.2,cex.names=1.2,
  main ="Most frequent words",
  ylab = "Word frequencies",col = terrain.colors(20))
termFrequency<- rowSums(as.matrix(dtm))
termFrequency<- subset(termFrequency,termFrequency>=48)
text(k,sort(termFrequency, decreasing = T)-
  1,labels=sort(termFrequency, decreasing = T),pch = 6, cex = 1)

```

Lampiran 5. Output SVM

SVM – Kernel Linear

```

> models <- train_models(container,"svm",kernel='linear')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2506:3131]),
+       as.character(results[, "SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	242	12
positiftraining	44	328

```

> recall_accuracy(sentiment_all[2506:3131], results[, "SVM_LABEL"])
[1] 0.9105431
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.85	0.95	0.90
positiftraining	0.96	0.88	0.92

SVM- Kernel Radial

```

> models <- train_models(container,"svm",kernel='radial')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2506:3131]),
+       as.character(results[, "SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	243	11
positiftraining	70	302

```

> recall_accuracy(sentiment_all[2506:3131], results[, "SVM_LABEL"])
[1] 0.870607
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.78	0.96	0.86
positiftraining	0.96	0.81	0.88

SVM- Kernel Polynomial

```

> models <- train_models(container,"SVM",kernel='polynomial')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2506:3131]),
+       as.character(results[,"SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	125	129
positiftraining	3	369

```

> recall_accuracy(sentiment_all[2506:3131], results[,"SVM_LABEL"])
[1] 0.7891374
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.98	0.49	0.65
positiftraining	0.74	0.99	0.85

SVM- Kernel Sigmoid

```

> models <- train_models(container,"SVM",kernel='sigmoid')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2506:3131]),
+       as.character(results[,"SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	251	3
positiftraining	71	301

```

> recall_accuracy(sentiment_all[2506:3131], results[,"SVM_LABEL"])
[1] 0.8817891
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.78	0.99	0.87
positiftraining	0.99	0.81	0.89

50:50

```

> models <- train_models(container,"SVM",kernel='linear')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[1567:3131]),
+       as.character(results[,"SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	597	39
positiftraining	115	814

```

> recall_accuracy(sentiment_all[1567:3131], results[,"SVM_LABEL"])
[1] 0.9015974
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.84	0.94	0.89
positiftraining	0.95	0.88	0.91

60:40

```
> models <- train_models(container,"svm",kernel='linear')
> results <- classify_models(container, models)
> recall_accuracy(sentiment_all[1879:3131], results[, "SVM_LABEL"])
[1] 0.9034318
> table(as.character(sentiment_all[1879:3131]),
+       as.character(results[, "SVM_LABEL"]))
```

	negatiftraining	positiftraining
negatiftraining	488	21
positiftraining	100	644

```
> create_precisionRecallSummary(container, results)
      SVM_PRECISION SVM_RECALL SVM_FSCORE
negatiftraining    0.83    0.96    0.89
positiftraining    0.97    0.87    0.92
```

70:30

```
> models <- train_models(container,"svm",kernel='linear')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2192:3131]),
+       as.character(results[, "SVM_LABEL"]))
```

	negatiftraining	positiftraining
negatiftraining	366	16
positiftraining	89	469

```
> recall_accuracy(sentiment_all[2192:3131], results[, "SVM_LABEL"])
[1] 0.8882979
> create_precisionRecallSummary(container, results)
      SVM_PRECISION SVM_RECALL SVM_FSCORE
negatiftraining    0.80    0.96    0.87
positiftraining    0.97    0.84    0.90
```

80:20

```
> models <- train_models(container,"svm",kernel='linear')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2506:3131]),
+       as.character(results[, "SVM_LABEL"]))
```

	negatiftraining	positiftraining
negatiftraining	242	12
positiftraining	44	328

```
> recall_accuracy(sentiment_all[2506:3131], results[, "SVM_LABEL"])
[1] 0.9105431
> create_precisionRecallSummary(container, results)
      SVM_PRECISION SVM_RECALL SVM_FSCORE
negatiftraining    0.85    0.95    0.90
positiftraining    0.96    0.88    0.92
```

90:10

```

> models <- train_models(container,"svm",kernel='linear')
> results <- classify_models(container, models)
> table(as.character(sentiment_all[2819:3131]),
+       as.character(results[,"SVM_LABEL"]))

```

	negatiftraining	positiftraining
negatiftraining	124	3
positiftraining	13	173

```

> recall_accuracy(sentiment_all[2819:3131], results[,"SVM_LABEL"])
[1] 0.9488818
> create_precisionRecallSummary(container, results)

```

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
negatiftraining	0.91	0.98	0.94
positiftraining	0.98	0.93	0.95

