

# Analisis Sentimen Menerapkan *Lexicon-Learning Based* Untuk Melihat Opini Masyarakat Mengenai Protokol Kesehatan Dan Perkembangan Vaksin Covid-19 Di Indonesia Menggunakan *Dataset Twitter*

Jefry Sunupurwa Asri<sup>1</sup>, Sawali Wahyu<sup>2</sup>

<sup>1,2</sup> Fakultas Ilmu Komputer, Universitas Esa Unggul

Email Correspondent Author : [jefry.sunupurwa@esaunggul.ac.id](mailto:jefry.sunupurwa@esaunggul.ac.id)<sup>1</sup>, [sawaliwahyu@esaunggul.ac.id](mailto:sawaliwahyu@esaunggul.ac.id)<sup>2</sup>

**Abstract** — The current COVID-19 pandemic has caused major impacts and changes in various fields of life. The current use of health protocols is carried out with the 5M movement, namely wearing masks, washing hands, keeping distance, limiting and limiting mobility. The development of covid cases has entered a new phase, namely the vaccine, the vaccine is useful to trigger the body's immunity to fight the corona virus. Public opinion regarding health protocols and vaccines on social media has positive and negative impacts, especially in social media activities on the twitter application, the application is used to obtain data because of the accuracy level of the truth of opinion sentences (tweets) uploaded to twitter, to find out How do people think about a topic? The use of the lexicon-based approach is a sentiment analysis approach model that is often used in a study to determine classification, with a language dictionary or corpus equipped with a weight for each word as a language or lexical source. The results of this study were carried out by taking samples of twitter data of approximately 15000 tweets with 3 datasets that would result in a classification of positive, negative and neutral sentiments towards public views regarding health protocols and vaccine developments in Indonesia.

**Keyword** — *Lexicon-Based Approach, Sentiment Analysis, Vaccine, Health Protocol*

**Abstrak** — Pandemi covid-19 yang terjadi pada saat ini telah menimbulkan dampak dan perubahan besar dalam berbagai bidang kehidupan, Penggunaan protokol kesehatan di indonesia saat ini telah dilakukan dengan gerakan 5M yaitu memakai masker, mencuci tangan, menjaga jarak, menjauhi kerumunan dan membatasi mobilisasi. Perkembangan kasus covid telah memasuki babak baru yaitu tersedianya vaksin, vaksin bermanfaat untuk memicu imunitas tubuh untuk melawan virus corona. Opini masyarakat mengenai protokol kesehatan dan vaksin di *social media* memiliki dampak positif dan negatif, terlebih lagi dalam aktivitas media *social media* pada aplikasi *twitter*, aplikasi tersebut digunakan untuk memperoleh data karena tingkat akurasi dari kebenaran kalimat opini (*tweet*) yang diunggah ke twitter, untuk mengetahui bagaimana pendapat masyarakat terhadap suatu topik. Penggunaan metode *lexicon-based approach* adalah model pendekatan analisis sentimen yang sering digunakan dalam suatu penelitian untuk menentukan klasifikasi, dengan menggunakan sebuah kamus kata bahasa atau *corpus* yang dilengkapi dengan bobot pada setiap katanya sebagai sumber bahasa atau leksikal. Hasil penelitian ini dilakukan dengan mengambil sampel data twitter sebanyak kurang lebih 15000 tweet dengan 3 dataset yang akan menghasilkan klasifikasi sentimen positif, negatif dan netral terhadap pandangan masyarakat mengenai protokol kesehatan dan perkembangan vaksin di indonesia.

**Kata Kunci** — *Lexicon-Based Approach, Analisis Sentimen, Vaksin, Protokol Kesehatan*

## I. PENDAHULUAN

Pandemi COVID-19 merupakan musibah yang memilukan seluruh penduduk bumi. Seluruh segmen kehidupan manusia di bumi terganggu, tanpa kecuali pendidikan. Banyak negara memutuskan menutup sekolah, perguruan tinggi maupun universitas, termasuk Indonesia. Krisis benar-benar datang tiba-tiba, pemerintah di belahan bumi manapun termasuk Indonesia harus mengambil keputusan untuk menanggulangi virus covid-19 [1].

Berkembangnya kasus covid di Indonesia telah berlangsung selama 2 tahun, dibarengi dengan ketersediaan vaksin diindonesia. Terdapat sejumlah merek vaksin yang tersedia diindonesia dalam rangka untuk mengurangi kasus covid, ditingkat masyarakat menjadi pro dan kontra terkait pelaksanaan vaksinasi diindonesia, apakah vaksinasi untuk masyarakat merupakan hak atau kewajiban [2].

Saat ini pemerintah indonesia telah memulai proses vaksinasi pada awal tahun 2021 yaitu kepada Nakes, Petugas Publik, Tenaga Pendidik dan Lansia. Dengan dilakukannya tahap vaksinasi tersebut akan memunculkan herd immunity kepada penduduk, sehingga virus tersebut dapat dinetralisir penyebarannya. Dengan adanya perkembangan kasus covid dan vaksin diindonesia, memiliki dampak yang tidak baik di social media. Sebagian orang lebih suka menyampaikan pendapat (opini) melalui dunia maya (internet). Baik itu lewat blog, forum online, media sosial (seperti facebook, twitter atau instagram) [3].

Twitter merupakan salah satu media sosial yang memiliki pengguna perharinya 126 juta. Selain itu twitter memiliki Twitter Trending Topic yang merupakan daftar topik terpopuler dan diupdate setiap waktu. Selanjutnya kelebihan lain dari twitter semua postingan yang ada di Twitter dapat diakses oleh pengguna lainnya [4].

Dengan demikian ini, pada penelitian ini akan menghasilkan analisis sentimen terhadap opini masyarakat mengenai protokol kesehatan dan perkembangan vaksin covid-19 di Indonesia, dengan mengambil sampel data twitter sebanyak kurang lebih 15000 tweet dengan 3 dataset yang akan menghasilkan klasifikasi sentimen positif, negatif dan netral.

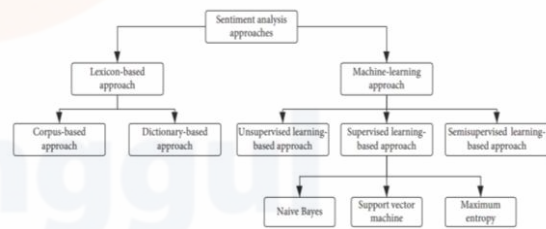
Analisis sentimen merupakan suatu proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, biasanya diklasifikasikan menjadi opini negatif dan positif. Analisis sentimen pada suatu kalimat menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu [5].

Pada pembahasan sebelumnya, twitter akan menjadi media yang memiliki kelebihan dibanding media sosial lainnya yaitu tampilan yang simple, topik terupdate, cepat dalam mengabarkan sebuah opini pengguna dan terbuka untuk mengakses tweet. Tentu ini semakin memudahkan dalam pengumpulan data opini dari masyarakat dan sebagai media penelitian [4]. Penggunaan *InSet Lexicon* digunakan untuk analisis sentimen data berbahasa Indonesia. *InSet Lexicon* (Indonesia Sentiment lexicon) terdiri dari 3.609 kata positif dan 6.609 kata negatif berbahasa Indonesia yang telah memiliki bobot nilai atau *polarity score* pada setiap katanya dengan kisaran bobot antara -5 sampai +5. *Polarity score* ini digunakan untuk mengklasifikasikan jenis sentimen. *InSet lexicon* juga dibangun untuk mengidentifikasi opini tertulis dan mengkategorikannya menjadi opini positif atau negatif yang bisa digunakan untuk menganalisis sentimen publik terhadap topik, acara, atau produk tertentu [6].

Pada penelitian ini menggunakan 2 pendekatan yaitu secara Learning Based (pendekatan dengan Machine Learning) dan Lexicon-Based (pendekatan berbasis leksikal). Pendekatan learning based akan menggunakan dataset yang telah diklasifikasikan secara dengan menyediakan dataset untuk menghasilkan klasifikasi text pendapat berdasarkan data twitter, sedangkan pendekatan lexicon-based akan digunakan sebagai kamus opini untuk penentuan klasifikasi [6]. Hasil penelitian ini akan menggabungkan 2 pendekatan tersebut untuk menghasilkan klasifikasi sentimen positif, negatif dan netral. Dengan 3 dataset yaitu opini masyarakat mengenai “Perkembangan Covid-19 di Indonesia”, “Pendapat Protokol Kesehatan”, dan “Pendapat Tentang Vaksin Covid”.

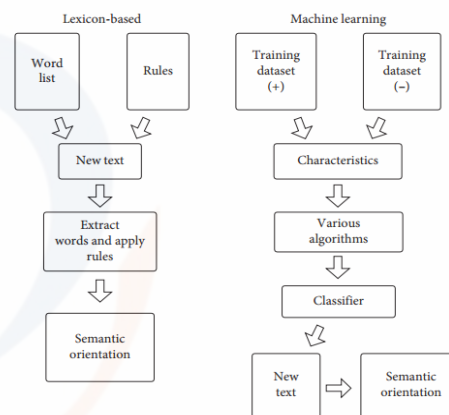
## II. TINJAUAN PUSTAKA

*Learning Based* (pendekatan dengan *Machine Learning*) dan *Lexicon-Based* (pendekatan berbasis leksikal). Pendekatan learning based akan menggunakan dataset yang telah diklasifikasikan secara dengan menyediakan dataset untuk menghasilkan klasifikasi text pendapat berdasarkan data twitter. Dengan kata lain, ketika ada teks telah dipilih, kata-kata di dalam teks harus cocokkan dengan kata-kata di kamus sentimen lalu berbagai algoritma telah digunakan untuk menggabungkan nilai-nilai. total agregasi nilai positif dan negatif dari kata-kata mengumpulkan orientasi sentimen dari keseluruhan teks [7].



Gambar I Sentimen Analysis Approachs

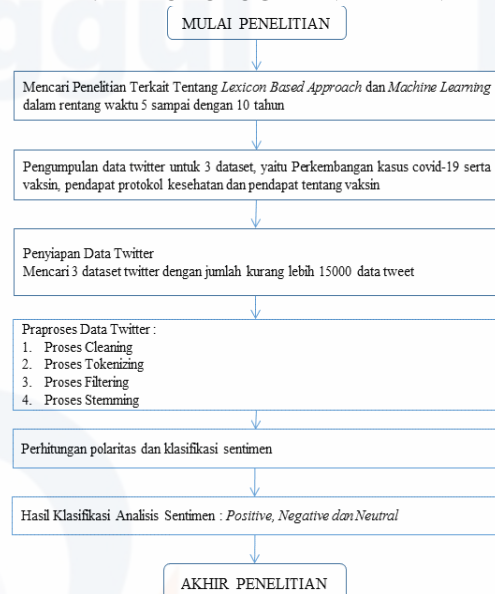
Pendekatan berbasis leksikon, bagaimanapun, memiliki keterbatasannya sendiri. Terkadang, ekstraksi kata individual mungkin terlewatkan makna penting dalam teks. Sejak sentimen yang ada kamus tidak dapat memenuhi konteks spesifik surat untuk pemegang saham, misalnya, dalam kamus, kata-kata seperti lebih rendah dan penurunan sama sekali tidak memiliki arti [7].



Gambar II Ilustration Lexicon Based With Machine Learning

Pada penelitian sebelumnya, penggunaan Lexicon Approach ini mampu meningkatkan hasil nilai precision dan nilai recall pada lexicon base [3].

## III. METODOLOGI PENELITIAN



Gambar III Langkah Penelitian

Penjelasan Gambar III.

- 1) Mencari Penelitian Terkait Tentang *Lexicon Based Approach* dan *Machine Learning* dalam rentang waktu 5 sampai dengan 10 tahun.
- 2) Pengumpulan data twitter untuk 3 dataset, yaitu Perkembangan kasus covid-19 serta vaksin, pendapat protokol kesehatan dan pendapat tentang vaksin
- 3) Penyiapan Data Twitter  
Yaitu Mencari 3 dataset twitter dengan jumlah kurang lebih 15000 data tweet, masing masing dataset terdiri dari 5000 data tweet.
- 4) Praproses Data Twitter :
  - a) Proses Cleaning yaitu dilakukan dengan membersihkan data tweet yang masih belum diolah / mentah dari karakter atau elemen yang tidak diperlukan.
  - b) Proses Tokenizing yaitu dilakukan dengan memecah nilai string dalam kalimat tweet menjadi satuan kata yang menyusun.
  - c) Proses Filtering yaitu dilakukan dengan pengambilan kata kata penting.
  - d) Proses Stemming yaitu dilakukan dengan mengubah kata dasar (stem) dengan cara menghilangkan imbuhan pada kata.
- 5) Perhitungan polaritas dan klasifikasi sentiment dilakukan dengan Proses perhitungan *polarity score* dilakukan dengan cara menjumlahkan keseluruhan bobot dari kata yang terdeteksi oleh sistem dan kemudian data tweet akan diklasifikasikan ke dalam jenis sentimen melalui algoritma yang diterapkan.
- 6) Hasil Klasifikasi Analisis Sentimen : *Positive, Negative dan Neutral.*

#### IV. HASIL DAN PEMBAHASAN

##### *Pengumpulan Data*

Dataset yang digunakan dalam penelitian ini berasal dari Twitter. Twitter dipilih sebagai sumber data pada penelitian ini dikarenakan twitter menjadi media sosial yang cukup digemari masyarakat. Twitter lebih unggul sebagai penyalur info/berita tercepat yang berbeda dari media sosial lainnya, karena ketika ada topik baru yang muncul (trending) dan menarik perhatian maka akan ada banyak sekali opini masyarakat yang muncul di twitter. Opini ini bisa menjadi sumber data yang akurat untuk melakukan analisis sentimen. Pengambilan data twitter memanfaatkan Twitter API. Kata kunci yang digunakan adalah frasa yang menjadi topik dasar dalam penelitian ini yaitu “Perkembangan Covid-19 di Indonesia”, “Pendapat

Protokol Kesehatan”, dan “Pendapat Tentang Vaksin Covid”.

Dengan jumlah data awal yang diambil untuk dataset “Perkembangan Covid-19 di Indonesia” sebanyak 5000 tweet, untuk dataset “Prokes Corona Terhadap Vaksin” sebanyak kurang lebih 5000 tweet, dan untuk dataset “Pendapat Tentang Vaksin Covid” sebanyak kurang lebih 5000 tweet.

Pembatasan jumlah data awal dilakukan agar langkah selanjutnya dapat berjalan dengan baik karena jika terlalu banyak data justru akan memperlambat praproses data Kurang Lebih 15000 tweet dari 3 dataset yang akan menjadi jumlah yang cukup ideal untuk mendapat beragam opini.

##### *Pra-Proses Data*

Pada tahap ini akan dijelaskan langkah selanjutnya dari pengumpulan data twitter, untuk menghasilkan data yang bersih dan terstruktur sehingga mampu memberikan hasil klasifikasi yang lebih akurat. Berikut adalah tahapan praproses data yang akan dilakukan :

- a) Proses Cleaning yaitu dilakukan dengan membersihkan data tweet yang masih belum diolah / mentah dari karakter atau elemen yang tidak diperlukan. Diantaranya yaitu
    - a. Hastag Twitter (#)
    - b. Retweet
    - c. *@username* (*mentions* dengan *username* twitter).
    - d. Link URL
    - e. Simbol, Angka dan tanda baca pada tweet
- Tahap tersebut dilakukan ke 3 dataset yang telah dilakukan pengambilan data kurang lebih 15000 data tweet. Berikut adalah sample data tweet yang dilakukan cleaning :

TABEL 1 CLEANING DATASET PERTAMA

Nama Dataset	Perkembangan Covid-19 dan Vaksin	
Proses	Sebelum	Sesudah
Remove Punctuation	@rurizrizal : Payah. Kalah sama indonesia yang ga kenal lock down. Sama corona aja kok takut. Haha. Ga ada yang bisa ngalahin negara ini lah pokoknya. Siapa dulu presidennya ?	Payah. Kalah sama indonesia yang ga kenal loc kdown. Sama corona aja kok takut. Haha. Ga ada yang bisa ngalahin negara ini lah pokoknya. Siapa dulu presidennya
Case Folding	Payah. Kalah sama indonesia yang ga kenal lock down. Sama corona aja kok takut. Haha. Ga ada yang bisa ngalahin negara ini lah pokoknya. Siapa dulu presidennya	payah. kalah sama indonesia yang ga kenal lock down. sama corona aja kok takut. haha. ga ada yang bisa ngalahin negara ini lah pokoknya. siapa dulu presidennya
Spelling Correction	Payah. Kalah sama indonesia yang ga kenal lock down. Sama corona aja kok takut. Haha. Ga ada yang bisa ngalahin negara ini lah pokoknya. Siapa dulu presidennya	Payah, kalah sama Indonesia yang tidak kenal lock down. Sama corona saja takut. Haha. Tidak ada yang bisa mengalahkan negara milah pokoknya. Siapa dulu presidennya

TABEL 2 CLEANING DATASET KEDUA

Nama Dataset	Protokol Kesehatan	
Proses	Sebelum	Sesudah
Remove Punctuation	@Txdarkebaikan: YA ALLAH ... SEMOGA ALLAH SEGERA SEMBUHKAN ... YUK TERUS SEMANGAT JAGA PROTOKOL KESEHATAN ... SALAH SATU KORBAN COVID ...	YA ALLAH SEMOGA ALLAH SEGERA SEMBUHKAN, YUK TERUS SEMANGAT JAGA PROTOKOL KESEHATAN, SALAH SATU KORBAN COVID
Case Folding	YA ALLAH SEMOGA ALLAH SEGERA SEMBUHKAN, YUK TERUS SEMANGAT JAGA PROTOKOL KESEHATAN, SALAH SATU KORBAN COVID	ya allah semoga allah segera sembuahkan, yuk terus semangat jaga protokol kesehatan, salah satu korban covid
Spelling Correction	ya allah semoga allah segera sembuahkan, yuk terus semangat jaga protokol kesehatan, salah satu korban covid	Ya allah, semoga allah segera sembuahkan, ayoo terus semangat jaga protokol kesehatan, salah satu korban covid.

TABEL 3 CLEANING DATASET KETIGA

Nama Dataset	Vaksin Covid-19	
Proses	Sebelum	Sesudah
Remove Punctuation	@RajaGuguk14 : nggak masalah yg saya curiga orang2 di indonesia setelah vaksin kedua tdk ada efek spt flu, ini yg membuat saya curiga. badan masih bersih+ vaksin pertama= membentuk kekebalan. vaksin kedua =antibodi/kekebalan sudah terbentuk membuat perlawanan vaksin ke2->spt sakit flu, pusing2	nggak masalah, yg saya curiga orang2 di indonesia setelah vaksin kedua tdk ada efek spt flu, ini yg membuat saya curiga. badan masih bersih, vaksin pertama : membentuk kekebalan. vaksin kedua : antibodi/kekebalan sudah terbentuk membuat perlawanan vaksin ke 2 spt sakit flu, pusing
Case Folding	nggak masalah, yg saya curiga orang di indonesia setelah vaksin kedua tdk ada efek spt flu, ini yg membuat saya curiga. badan masih bersih, vaksin pertama : membentuk kekebalan. vaksin kedua : antibodi/kekebalan sudah terbentuk membuat perlawanan vaksin ke 2 spt sakit flu, pusing	nggak masalah, yg saya curiga orang2 di indonesia setelah vaksin kedua tdk ada efek spt flu, ini yg membuat saya curiga. badan masih bersih, vaksin pertama : membentuk kekebalan. vaksin kedua : antibodi/kekebalan sudah terbentuk membuat perlawanan vaksin ke 2 spt sakit flu, pusing
Spelling Correction	nggak masalah, yg saya curiga orang di indonesia setelah vaksin kedua tdk ada efek spt flu, ini yg membuat saya curiga. badan masih bersih, vaksin pertama : membentuk kekebalan. vaksin kedua : antibodi/kekebalan sudah terbentuk membuat perlawanan vaksin ke2 spt sakit flu, pusing2	Tidak masalah. Yang saya curiga orang di Indonesia setelah Vaksin kedua tidak ada efek seperti flu, ini yang membuat saya curiga. Badan masih bersih, vaksin pertama membentuk kekebalan. Vaksin kedua antibody/kekebalan sudah terbentuk membuat perlawanan vaksin ke 2 seperti sakit flu, pusing - pusing.

- b) Proses Tokenizing yaitu dilakukan dengan memecah nilai string dalam kalimat tweet menjadi satuan kata yang menyusun. Proses ini dilakukan untuk pemenggalan kalimat menjadi kata. Berikut adalah proses tokenizing :

TABEL 4 TOKENIZING DATASET PERTAMA

Nama Dataset	Perkembangan Covid-19 dan Vaksin
Sebelum	Sesudah
Payah, kalah sama Indonesia yang tidak kenal lock down. Sama corona saja takut. Haha. Tidak ada yang bisa mengalahkan negara inilah pokoknya. Siapa dulu presidennya	(“Payah”, “kalah”, “sama”, “Indonesia”, “yang”, “tidak”, “kenal”, “lockdown”, “Sama”, “corona”, “saja”, “takut”, “Haha”, “Tidak”, “ada”, “yang”, “bisa”, “mengalahkan”, “negara”, “inilah”, “pokoknya”, “Siapa”, “dulu”, “presidennya”).

TABEL 5 TOKENIZING DATASET KEDUA

Nama Dataset	Protokol Kesehatan
Sebelum	Sesudah
Ya allah, semoga allah segera sembuhkan, ayo terus semangat jaga protocol kesehatan, salah satu korban covid.	(“Ya”, “allah”, “semoga”, “allah”, “segera”, “sembuhkan”, “ayo”, “terus”, “semangat”, “jaga”, “protokol”, “kesehatan”, “salah”, “satu”, “korban”, “covid”).

TABEL 6 TOKENIZING DATASET KETIGA

Nama Dataset	Vaksin Covid-19
Sebelum	Sesudah
Tidak masalah. Yang saya curiga orang di Indonesia setelah Vaksin kedua tidak ada efek seperti flu, ini yang membuat saya curiga. Badan masih bersih, vaksin pertama membentuk kekebalan. Vaksin kedua antibody/kekebalan sudah terbentuk membuat perlawanan vaksin ke 2 seperti sakit flu, pusing - pusing	(“Tidak”, “masalah”, “Yang”, “saya”, “curiga”, “orang”, “di Indonesia”, “setelah”, “Vaksin”, “kedua”, “tidak”, “ada”, “efek”, “seperti”, “flu”, “ini”, “yang”, “membuat”, “saya”, “curiga”, “Badan”, “masih”, “bersih”, “vaksin”, “pertama”, “membentuk”, “kekebalan”, “Vaksin”, “kedua”, “antibody”, “kekebalan”, “sudah”, “terbentuk”, “perlawanan”, “vaksin”, “ke”, “2”, “seperti”, “sakit flu”, “pusing - pusing”).

- c) Proses Filtering yaitu dilakukan dengan pengambilan kata kata penting. Penggunaan stopword atau kata yang kurang memiliki makna akan dihilangkan karena tidak diperlukan.

TABEL 7. FILTERING DATASET PERTAMA

Nama Dataset	Perkembangan Covid-19 dan Vaksin
Sebelum	Sesudah
(“Payah”, “kalah”, “sama”, “Indonesia”, “yang”, “tidak”, “kenal”, “lockdown”, “Sama”,	(‘lockdown’, ‘corona’, ‘mengalahkan’, ‘negara’, ‘presiden’)

“corona”, “saja”, “takut”, “Haha”, “Tidak”, “ada”, “yang”, “bisa”, “mengalahkan”, “negara”, “inilah”, “pokoknya”, “Siapa”, “dulu”, “pres idennya”).
---

TABEL 8 FILTERING DATASET KEDUA

Nama Dataset	Protokol Kesehatan
Sebelum	Sesudah
(“Ya”, “allah”, “semoga”, “allah”, “segera”, “sembuhkan”, “ayo”, “terus”, “semangat”, “jaga”, “protokol”, “kesehatan”, “salah”, “satu”, “korban”, “covid”).	(‘allah’, ‘sembuhkan’, ‘semangat’, ‘protokol’, ‘kesehatan’, ‘korban’, ‘covid’).

TABEL 9 FILTERING DATASET KETIGA

Nama Dataset	Vaksin Covid-19
Sebelum	Sesudah
(“Tidak”, “masalah”, “Yang”, “saya”, “curiga”, “orang”, “di Indonesia”, “setelah”, “Vaksin”, “kedua”, “tidak”, “ada”, “efek”, “seperti”, “flu”, “ini”, “yang”, “membuat”, “saya”, “curiga”, “Badan”, “masih”, “bersih”, “vaksin”, “pertama”, “membentuk”, “kekebalan”, “Vaksin”, “kedua”, “antibody”, “kekebalan”, “sudah”, “terbentuk”, “perlawanan”, “vaksin”, “ke”, “2”, “seperti”, “sakit flu”, “pusing - pusing”).	(‘curiga’, ‘indonesia’, ‘vaksin’, ‘efek’, ‘flu’, ‘membuat’, ‘badan’, ‘bersih’, ‘antibodi’, ‘pusing-pusing’).

- d) Proses Stemming yaitu dilakukan dengan mengubah kata dasar (stem) dengan cara menghilangkan imbuhan pada kata berupa awalan maupun akhiran.

TABEL 10 STEMMING DATASET PERTAMA

Nama Dataset	Perkembangan Covid-19 dan Vaksin
Sebelum	Sesudah
(‘lockdown’, ‘corona’, ‘mengalahkan’, ‘negara’, ‘presiden’)	(‘lockdown’, ‘corona’, ‘negara’, ‘presiden’)

TABEL 11 STEEMING DATASET KEDUA

Nama Dataset	Protokol Kesehatan
Sebelum	Sesudah
(“Ya”, “allah”, “semoga”, “allah”, “segera”, “sembuhkan”, “ayo”, “terus”, “semangat”, “jaga”, “protokol”, “kesehatan”, “salah”, “satu”, “korban”, “covid”).	(‘allah’, ‘sembuhkan’, ‘semangat’, ‘protokol’, ‘kesehatan’, ‘korban’, ‘covid’).

TABEL 12 STEMMING DATASET KETIGA

Nama Dataset	Vaksin Covid-19
Sebelum	Sesudah
(‘curiga’, ‘indonesia’, ‘vaksin’, ‘efek’, ‘flu’, ‘membuat’, ‘badan’,	(‘indonesia’, ‘vaksin’, ‘efek’, ‘flu’, ‘badan’,

'bersih', 'antibodi', 'pusing-pusing').	'antibodi', 'pusing-pusing').
---	-------------------------------

menggunakan algoritma dengan konsep InSet Lexicon sebagai kamus opini dari data yang diambil pada tweet. Berikut adalah hasil perhitungan polarity score :

**Hasil Polarity Score**

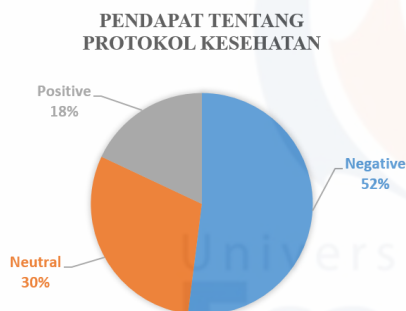
Bagian ini menjelaskan hasil polarity score dari data hasil preproses data yang sudah diolah sebanyak 15000 data, kemudian diklasifikasikan secara otomatis

TABEL 13 HASIL POLARITY SCORE DATASET

Nama DataSet	Tweet Preprocessed	Score	Presentase	Polarity	Total Presentase
Protokol Kesehatan	('protokol', 'kesehatan', 'covid19', 'mematuhi', 'masyarakat', 'mengingatkan', 'penyebaran').	2.154	52%	Negative	100%
		1.243	30%	Neutral	
		746	18%	Positive	
Vaksin Covid-19	('vaksin', 'covid19', 'sinovac', 'harapan', 'dosis', 'kedatangan', 'pandemi').	1.842	30%	Negative	100%
		1.535	25%	Neutral	
		2.763	45%	Positive	
Perkembangan Covid-19 dan Vaksin	('varian', 'vaksin', 'corona', 'covid19', 'sudah', 'ada', 'prokes').	1689	43%	Negative	100%
		982	25%	Neutral	
		1.257	32%	Positive	

Berdasarkan tabel 13 dijelaskan data tweet yang telah diklasifikasikan ke dalam kelasnya berdasarkan polarity score dari tiap kata yang digunakan InSet Lexicon. Berikut adalah hasil presentase pemetaan klasifikasi yang ditampilkan dalam bentuk diagram chart dengan keterangan presentase :

**Hasil Klasifikasi pada Dataset Protokol Kesehatan**

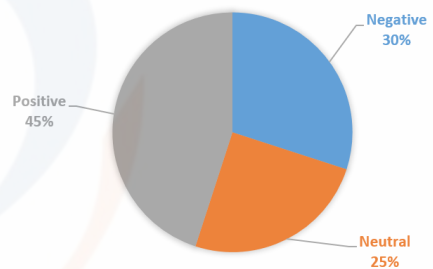


Gambar III. Hasil Klasifikasi Protokol Kesehatan

Pada gambar 3 dijelaskan, hasil klasifikasi pada dataset protokol kesehatan, memiliki presentase diantaranya : 52% Negative, 30% Neutral, dan 18% Positive. Dengan demikian opini masyarakat terhadap dataset protokol kesehatan yaitu tertinggi adalah pandangan negative, artinya tweet yang didapatkan memiliki sifat yang tidak baik.

**Hasil Klasifikasi pada Dataset Vaksin Covid-19**

PENDAPAT TENTANG VAKSIN

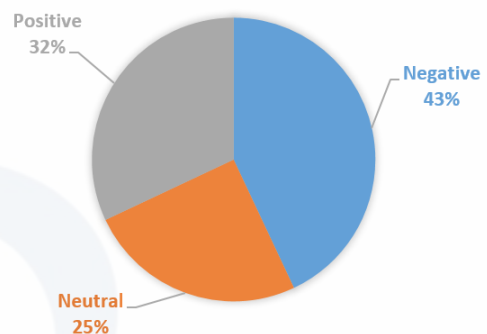


Gambar IV. Hasil Klasifikasi Vaksin Covid-19

Pada gambar 4 dijelaskan, hasil klasifikasi pada dataset vaksin covid-19, memiliki presentase diantaranya : 30% Negative, 25% Neutral, dan 45% Positive. Dengan demikian opini masyarakat terhadap dataset vaksin covid-19 yaitu tertinggi adalah pandangan positive, artinya tweet yang didapatkan memiliki pandangan positive terhadap vaksin yang telah berjalan di Indonesia.

**Hasil Klasifikasi pada Dataset Perkembangan Covid-19 dan Vaksin**

PENDAPAT TENTANG PERKEMBANGAN COVID-19 DAN VAKSIN



Gambar V. Hasil Klasifikasi Perkembangan Covid-19 dan Vaksin

Pada gambar 5 dijelaskan, hasil klasifikasi pada dataset perkembangan covid-19 dan vaksin covid-19, memiliki presentase diantaranya : 43% Negative, 25% Neutral, dan 32% Positive. Dengan demikian opini masyarakat terhadap dataset perkembangan kasus covid-19 dan vaksin covid-19 yaitu tertinggi adalah pandangan negative, artinya tweet yang didapatkan memiliki pandangan negatif terhadap perkembangan kasus covid-19 dan vaksin di Indonesia.

Visualisasi dari kata-kata yang muncul pada keseluruhan data tweet kemudian ditampilkan dalam bentuk Word Cloud.

Word cloud akan mempermudah penggambaran kata dengan memberikan visualisasi data teks yang menarik dan informatif. Frekuensi kata dalam keseluruhan data mempengaruhi besar atau kecilnya kata tersebut pada word cloud. Semakin tinggi frekuensi kata maka ukuran kata tersebut pada word cloud akan semakin besar, begitu pula sebaliknya. Kata dengan frekuensi yang sedikit akan ditampilkan dengan ukuran yang lebih kecil. Berikut adalah hasil word cloud yang telah didapatkan :

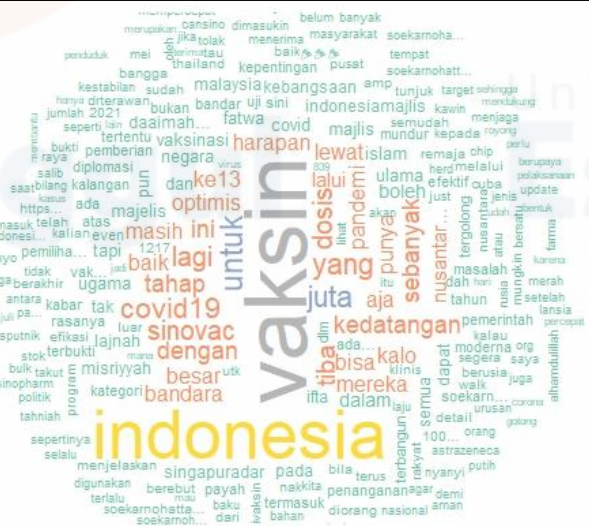
**Word Cloud pada Dataset Protokol Kesehatan**



Gambar VI. Word Cloud Protokol Kesehatan

Gambar VI menjelaskan hasil word cloud pada protocol kesehatan yang memiliki frekuensi besar adalah kata protocol kesehatan, meningkatkan, melakukan, mematuhi dan lain lain.

**Word Cloud pada Dataset Vaksin Covid-19**



Gambar VII Word Cloud Vaksin Covid-19

Gambar VII menjelaskan hasil word cloud pada protocol kesehatan yang memiliki frekuensi besar adalah kata Indonesia, vaksin, harapan, dosis dan lain lain.

**Word Cloud pada Dataset Perkembangan Covid-19 dan Vaksin**



Gambar VIII. Word Cloud Perkembangan Covid-19 dan Vaksin

Gambar VIII menjelaskan hasil word cloud pada protocol kesehatan yang memiliki frekuensi besar adalah kata Indonesia, corona, prokes, varian, vaksin dan lain lain.

**V. KESIMPULAN**

Pada bagian ini dijelaskan hasil kesimpulan dari penelitian yang telah dilakukan adalah sebagai berikut :

- 1) Penelitian ini berhasil menggabungkan 2 pendekatan yaitu dengan Learning Based (pendekatan dengan Machine Learning) dan Lexicon-Based (pendekatan berbasis leksikal);
- 2) Penelitian ini Menghasilkan klasifikasi sentimen positif, negatif dan netral. Dengan 3 dataset yaitu

opini masyarakat mengenai pendapat protokol kesehatan, vaksin covid-19, dan perkembangan kasus covid dan vaksin;

- 3) Penelitian ini berhasil mendapatkan klasifikasi pada dataset protokol kesehatan, dengan presentase diantaranya : 52% Negative, 30% Neutral, dan 18% Positive;
- 4) Penelitian ini berhasil mendapatkan klasifikasi pada dataset vaksin covid-19, memiliki presentase diantaranya : 30% Negative, 25% Neutral, dan 45% Positive;
- 5) Penelitian ini berhasil mendapatkan klasifikasi pada dataset perkembangan covid-19 dan vaksin covid-19, memiliki presentase diantaranya : 43% Negative, 25% Neutral, dan 32% Positive.

#### UCAPAN TERIMA KASIH

Ucapan Terimakasih diberikan kepada :

- 1) Kepada ketua penyelenggara Konferensi Nasional Ilmu Komputer (Konik) 2021 dan bapak/ibu editor serta reviewer yang telah memberikan kesempatan kepada penulis untuk dapat berpartisipasi dalam jurnal prosiding Konik 2021.
- 2) Kepada Bapak Sawali Wahyu, S.Kom, M.Kom selaku rekan satu penelitian yang telah membantu memberikan arahan, motivasi dan bimbingan dalam penelitian ini sampai selesai.

#### DAFTAR ACUAN

- [1] R. H. Syah, "Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran," *SALAM J. Sos. dan Budaya Syar-i*, vol. 7, no. 5, 2020, doi: 10.15408/sjsbs.v7i5.15314.
- [2] F. Gandryani and F. Hadi, "Pelaksanaan Vaksinasi Covid-19 Di Indonesia: Hak Atau Kewajiban Warga Negara (the Vaccination of Covid-19 in Indonesia: Citizen Right or Citizen Duty)," *J. Rechts Vinding Media Pemb. Huk. Nas.*, vol. 10, no. 1, p. 23, 2021, [Online]. Available: <https://setkab.go.id/gallery/cegah-Covid-19-dengan-3m->.
- [3] Y. Azhar, "Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 3, p. 237, 2018, doi: 10.23887/janapati.v6i3.11739.
- [4] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.
- [5] E. Mona Cindo, Dian Palupi Rini, "Literatur Review : Metode Klasifikasi Pada Sentimen Analisis," *Semin. Nas. Teknol. Komput. Sains*, pp. 66–70, 2019.
- [6] D. Musfiroh, U. Khaira, P. Eko, P. Utomo, and T. Suratno, "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 24–33, 2021.
- [7] S. Che, W. Zhu, and X. Li, "Anticipating Corporate Financial Performance from CEO Letters Utilizing Sentiment Analysis," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/5609272.