

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pandemi yang diakibatkan oleh virus COVID-19 adalah salah satu peristiwa paling berdampak besar pada abad ke-21. Efek dari pandemi kepada seluruh dunia sangatlah masif. Hampir seluruh sektor kehidupan terkena dampak dari adanya pandemi COVID-19. Berbagai macam usaha dilakukan oleh manusia untuk tetap bertahan pada keadaan yang sulit akibat dampak adanya pandemi COVID-19.

Salah satu usaha yang dilakukan untuk melewati masa sulit akibat pandemi COVID-19 adalah dengan membuat vaksin COVID-19. Adanya vaksin COVID-19 akan mampu mengubah hidup manusia di seluruh dunia menjadi sedia kala seperti sebelum adanya pandemi COVID-19.

Indonesia adalah salah satu negara yang bergerak cepat dalam melakukan usaha untuk mengembalikan hidup seperti sebelum adanya pandemi COVID-19 dengan memesan vaksin untuk COVID-19. Per 13 Desember tahun 2020, Indonesia telah memiliki beberapa vaksin, yaitu vaksin Bio Farma, vaksin Moderna, vaksin Pfizer, vaksin Sinopharm, dan vaksin Sinovac (Kompas, 2020).

Namun, pengadaan vaksin yang dilakukan oleh pemerintah menuai banyak pendapat pro dan kontra. Orang-orang yang berpendapat pro terhadap pengadaan vaksin COVID-19 yang dilakukan oleh pemerintah Indonesia di akhir 2020 ini pada umumnya berkaitan dengan kondisi ekonomi. Sedangkan orang-orang yang berpendapat kontra terhadap pengadaan vaksin COVID-19 yang dilakukan oleh pemerintah di akhir tahun 2020 pada umumnya berkaitan dengan belum selesainya uji klinis atas vaksin yang disediakan oleh pemerintah serta terkait biaya untuk mendapatkan vaksin tersebut.

Pemerintah sebagai pemangku kepentingan yang mengurus hajat hidup orang banyak harus dapat mengambil kebijakan dan keputusan berdasarkan suara rakyatnya. Oleh karena itu, diperlukan adanya penilaian persepsi publik terhadap vaksin COVID-19 di Indonesia. Melalui persepsi publik tersebut, dapat diketahui rekomendasi terkait COVID-19 yang mungkin dapat diambil oleh pemerintah agar sesuai dengan suara rakyat.

1.2 Rumusan Masalah

Adapun rumusan masalah umum berdasarkan latar belakang tersebut adalah:

1. Bagaimana persepsi publik terhadap vaksin COVID-19 di Indonesia?
2. Bagaimana kebijakan yang dapat diambil agar sesuai persepsi publik?

Beberapa turunan masalah spesifik dari rumusan masalah umum di atas adalah:

1. Bagaimana skor sentimen pendapat rakyat melalui *tweet* terkait vaksin COVID-19 di Indonesia?
2. Klasifikasi sentimen apa yang memiliki jumlah terbanyak?
3. Klasifikasi sentimen apa yang memiliki jumlah terkecil?
4. Bagaimana strategi kebijakan vaksin COVID-19 di Indonesia agar tepat guna dan tepat sasaran?

1.3 Pendekatan Penelitian

Pendekatan yang digunakan pada penelitian kali ini adalah dengan menggunakan *twitter* untuk menarik persepsi public dan menggunakan *software* RStudio dalam melakukan pengolahan data. Kemudian, pendekatan yang digunakan dalam menentukan persepsi publik adalah dengan melakukan analisis sentiment dengan menggunakan pendekatan leksikon.

1.4 Tujuan dan Manfaat Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut:

1. Mengetahui persepsi publik terkait vaksin COVID-19 di Indonesia.
2. Mengetahui klasifikasi sentimen yang memiliki jumlah paling banyak.
3. Mengetahui klasifikasi sentimen yang memiliki jumlah paling sedikit.
4. Mengetahui strategi kebijakan terkait vaksin COVID-19 yang tepat guna dan tepat sasaran.

BAB 2

DASAR TEORI

2.1 Text Mining

2.1.1 Definisi Text Mining

Text mining adalah salah satu bentuk variasi dari *data mining* yang dapat didefinisikan sebagai proses pencarian korelasi, pola, dan tren dengan menelaah sejumlah besar data yang disimpan dalam media penyimpanan tertentu menggunakan *teknologi pattern recognition* serta teknik matematis dan statistik (Larose, 2005). Sumber lain mengatakan bahwa *text mining* adalah sebuah proses mencari pengetahuan, identifikasi, dan eksplorasi pola yang menarik dari suatu dokumen yang dapat dilakukan menggunakan beberapa alat analisis (Feldman & Sanger, 2007). Kunci dari pelaksanaan *text mining* adalah mencari koneksi dari informasi yang telah diekstrak tersebut untuk membentuk fakta atau hipotesis baru yang akan digali lebih lanjut. Berdasarkan kedua definisi tersebut, dapat disimpulkan bahwa *text mining* adalah proses mengolah data tertulis dengan menggunakan teknik matematis dan statistik untuk menemukan pola tertentu agar didapat sebuah informasi baru.

Berbeda dari *data mining* yang hanya menggunakan data terstruktur, *text mining* dapat dilakukan menggunakan data yang terstruktur seperti basis data atau file XML maupun data yang tidak terstruktur atau semi-terstruktur, seperti surat elektronik, dokumen tertulis, maupun file HTML (Fan, Wallace, Rich, & Zhang, 2006). Maka dari itu, *text mining* banyak digunakan perusahaan saat ini untuk memperoleh solusi dan pengetahuan yang lebih baik mengingat jumlah data semi-terstruktur dan tidak terstruktur yang jauh lebih banyak dibanding dengan data terstruktur.

2.1.2 Teknik Text Mining

Menurut Fan, Wallace, Rich, & Zhang pada tahun 2006, terdapat berbagai macam teknik yang dapat digunakan dalam melakukan *text mining* yang akan dijelaskan sebagai berikut.

1. Ekstraksi informasi (*information extraction*)

Teknik ini bermaksud untuk menganalisis teks yang tidak terstruktur dan mengidentifikasi frasa kunci dan hubungannya dalam teks dengan cara mencari urutan yang telah ditetapkan dalam teks.

2. Pencarian topik (*topic tracking*)

Sistem ini menyimpan profil dari pengguna dan memprediksi dokumen yang mungkin membuat pengguna tersebut tertarik berdasarkan sejarah dokumen yang dilihat oleh pengguna.

3. *Summarization*

Teknologi ini bermaksud untuk membantu pengguna untuk merangkum dan mengurangi panjang dan detail dari sebuah dokumen tanpa mengubah makna dari keseluruhan dokumen tersebut.

4. *Categorization*

Categorization berfungsi untuk mengidentifikasi tema dari sebuah dokumen dengan cara menghitung kata-kata yang muncul dengan mempertimbangkan kata-kata yang terkait. Biasanya telah ada kategori tertentu yang ditentukan sebelumnya.

5. *Clustering*

Teknik ini berfungsi untuk mengelompokkan dokumen yang mirip, namun, berbeda dengan *categorization*, topik dari masing-masing kelompok belum ditentukan.

6. *Concept linkage*

Tujuan dari metode ini adalah menghubungkan dokumen yang terkait berdasarkan hasil identifikasi konsep yang digunakan, biasanya digunakan dalam bidang biomedis dan sejenisnya.

7. *Visualization*

Visualisasi informasi bermaksud untuk menyajikan bentuk teks dalam jumlah besar ke dalam bentuk pemetaan atau hierarki visual sehingga lebih mudah untuk dipahami.

8. *Question answering*

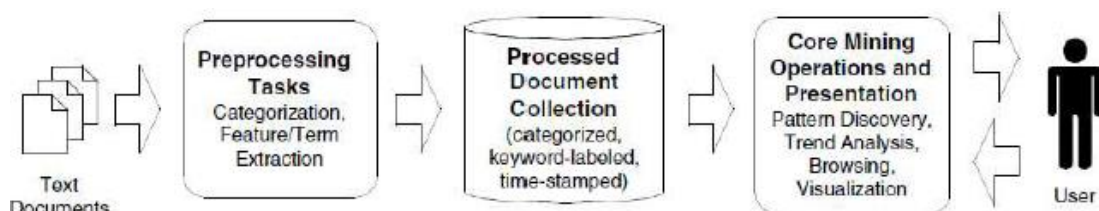
Hal ini bermaksud untuk memberikan jawaban terbaik yang mungkin untuk sebuah pertanyaan yang diberikan menggunakan teknik *text mining* lainnya, seperti ekstraksi informasi, kategorisasi, dan lain-lain.

9. Analisis sentimen (*sentiment analysis*)

Analisis sentimen atau *opinion mining* bertujuan untuk mengklasifikasikan ulasan atau review berdasarkan polaritasnya (negatif atau positif) (Pang & Lee, 2008).

2.1.3 Arsitektur Fungsional Text Mining

Pada implementasinya secara umum, sistem *text mining* mengambil masukan berupa dokumen mentah dan dapat menghasilkan berbagai macam tipe keluaran, seperti pola, pemetaan koneksi, dan tren (Feldman & Sanger, 2007). Dalam level fungsional, sistem text mining mengikuti model umum yang disediakan oleh beberapa aplikasi *data mining* klasik dan dibagi menjadi empat (4) area utama. Gambar 1 adalah penjelasan dari empat (4) area yang disebutkan sebelumnya.



Gambar 1 Arsitektur Fungsional Text Mining

Sumber: (Feldman & Sanger, 2007)

1. *Preprocessing Task*

Preprocessing task mencakup seluruh prosedur, proses, dan metode yang dibutuhkan dalam mempersiapkan data untuk operasi penemuan pengetahuan (*knowledge discovery*) dalam inti dari sistem text mining. Proses *data preprocessing* secara umum mengonversi informasi dari setiap sumber data menjadi bentuk matematis sehingga dapat diekstraksi untuk membuat dokumen baru yang lebih bermakna atau memiliki makna. Hal ini akan dibahas dalam subbab selanjutnya.

2. *Core Mining Operation*

Core mining operation adalah proses inti dalam suatu *text mining*. Pada proses ini, termasuk pencarian pola (*pattern discovery*), analisis tren, dan algoritma pencarian pengetahuan. Pola yang sering digunakan dalam pencarian pengetahuan teks adalah distribusi, asosiasi, atau perbandingan tingkat kemenarikan di antara pola data yang ada.

3. *Presentation Layer Components*

Pada area ini, komponen termasuk *graphic user interface* (GUI), fungsionalitas pencarian pola, serta akses terhadap *query*. Alat bantu visualisasi dan penyunting *query* untuk pengguna juga termasuk ke dalam kategori ini.

4. *Refinement Techniques*

Proses ini mencakup teknik untuk menyaring informasi yang redundan dan mengelompokkan data-data yang terhubung ke dalam sistem *text mining* yang telah ditentukan. Proses ini biasa disebut dengan *postprocessing techniques*.

2.2 Data Preprocessing Teks

Untuk mendapatkan seluruh kata yang digunakan dalam suatu teks tertentu, dibutuhkan adanya proses tokenisasi yang didefinisikan sebagai proses pemisahan dokumen teks menjadi aliran kata dengan cara menghapus seluruh tanda baca serta mengkonversi tabulasi dan karakter lain yang bukan dalam bentuk teks lainnya menjadi spasi (Feldman & Sanger, 2007). Kumpulan kata-kata berbeda yang didapat dari penyatuan seluruh dokumen teks disebut dengan kamus (*dictionary*) dari koleksi dokumen.

Untuk meminimasi ukuran dari kamus yang digunakan pada proses *text mining*, kumpulan kata yang menjelaskan dokumen tersebut dapat dikurangi dengan melakukan *proses filtering*, *metode lemmatization*, atau *stemming* (Feldman & Sanger, 2007).

1. Tokenisasi

Tokenisasi adalah proses pemisahan teks ke dalam bentuk kata tunggal yang lebih mudah untuk diproses. Proses pemecahan ini bisa terjadi pada beberapa tingkat aliran kata yang berbeda, seperti pada sebuah dokumen dapat dipecah menjadi paragraf, kalimat, kata, atau bahkan suku kata.

2. *Cleansing*

Tahap ini adalah pembersihan dokumen teks dari unsur yang tidak digunakan dalam proses *text mining* atau tidak memiliki pengaruh terhadap pemrosesan teks. Terdapat beberapa tahap dalam proses pembersihan, yaitu:

- a. Eliminasi simbol dan tanda baca
Simbol dan tanda baca umumnya tidak akan mengubah makna dari sebuah dokumen teks, sehingga lebih baik dieliminasi.
- b. *Case folding*
Penggunaan huruf kapital pada umumnya tidak mengubah makna dari sebuah dokumen teks dan pada saat pemrosesan dokumen teks, terdapat beberapa bahasa pemrograman yang sensitif terhadap perbedaan huruf kapital. Contohnya adalah kata “Bahagia” dan “bahagia” memiliki makna yang sama, namun pada beberapa bahasa pemrograman, kedua kata tersebut memiliki arti yang berbeda. Oleh karena itu, proses pengubahan seluruh huruf menjadi huruf kecil (*lowercase*) sangat penting untuk dilakukan.
- c. Eliminasi *stop words*
Stop words adalah kata-kata yang umum digunakan namun tidak memiliki nilai tambah dalam pengolahan data (Feinerer, 2008). Beberapa contoh *stop words* dalam bahasa Indonesia adalah “ada”, “di”, “yang”, “ke”, dan lain sebagainya. Eliminasi *stop words* dilakukan agar hasil pengolahan data lebih akurat.

3. *Stemming*

Stemming merupakan proses menghapus imbuhan pada suatu kata untuk mendapatkan kata dasar untuk mengurangi kompleksitas dari pengolahan data (Feinerer, 2008). Proses ini dilakukan untuk mengurangi variasi kata yang memiliki makna sama agar jumlah kata dalam bag of words dapat diminimasi. Contoh dari stemming yaitu mengubah kata ‘mengubah’ dan ‘diubah’ menjadi ‘ubah’. Pada analisis sentimen berbasis leksikon, proses *stemming* disesuaikan dengan kata yang terdapat pada kamus sentimen. Apabila kamus sentimen terdiri dari kata dasar, maka *stemming* perlu dilakukan. Sebaliknya, apabila kamus sentimen terdiri dari kata berimbuhan, maka proses *stemming* tidak perlu dilakukan.

4. Normalisasi kata

Jika dokumen data yang ditarik adalah dokumen teks yang berasal dari media sosial, maka akan banyak penggunaan kata yang tidak baku. Oleh karena itu, perlu dilakukan transformasi dari kata tersebut menjadi kata yang sesuai dengan KBBI (PUEBI).

2.3 Analisis Sentimen

Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisis opini, sentimen, evaluasi, pujian, sikap, dan emosi seseorang terhadap sebuah entitas seperti produk, jasa, atau organisasi, serta atributnya (Liu, 2012). Dalam praktiknya, suatu entitas ingin mengetahui opini atau perspektif terhadap entitas tersebut, termasuk suatu entitas negara. Analisis sentimen dapat dilakukan oleh suatu negara (melalui pemerintah) untuk menilai persepsi publik terhadap suatu isu atau kebijakan.

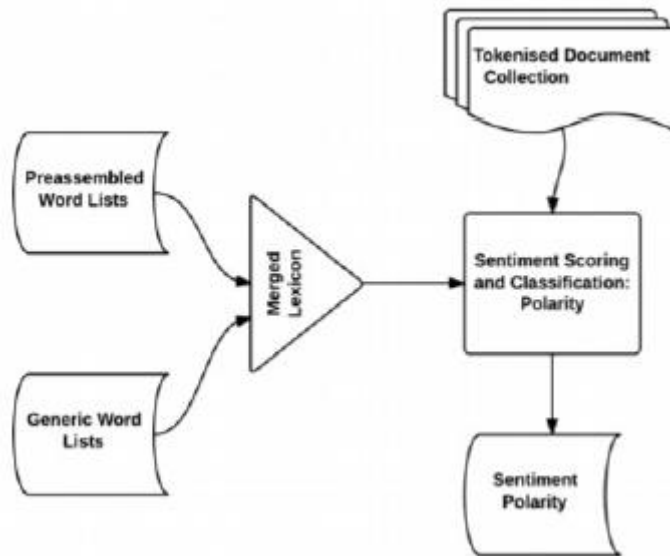
Dalam melakukan analisis sentimen, indikator utama yang digunakan adalah kata sentimen (*sentiment words*) atau kata opini (*opinion words*), yaitu kata-kata yang umumnya digunakan untuk mengungkapkan sentimen positif atau negatif (Liu, 2012). Misalnya, ‘bagus’, ‘suka’, dan ‘baik’ digunakan untuk mengekspresikan sentimen positif, sedangkan ‘jelek’, ‘benci’, dan ‘sulit’ digunakan untuk menggambarkan sentimen negatif. Sebuah data atau daftar yang berisi kata-kata opini ini disebut dengan *sentiment lexicon* atau *opinion lexicon*. Walaupun pada implementasinya, ada kesempatan di mana *sentiment lexicon* tidak dapat merepresentasikan sentimen dalam sebuah dokumen, terutama untuk dokumen yang kompleks.

Sentimen biasanya diklasifikasikan menjadi dua kelas, yaitu positif dan negatif, walaupun ada juga yang menambahkan kelas netral. Biasanya, dalam sebuah rating produk berskala 1 sampai 5, nilai 1 dan 2 dimasukkan ke dalam kelas ‘negatif’, nilai 3 menjadi kelas ‘netral’, sedangkan 4 dan 5 menjadi kelas ‘positif’.

Analisis sentimen dapat dikategorikan sebagai permasalahan klasifikasi teks, sehingga dapat diterapkan model klasifikasi yang lazim digunakan, seperti Naïve Bayes, *Support Vector Machines* (SVM), *Decision Tree*, dan lain sebagainya (Hotho, Nurnberger, & Neagu, 2005)

2.4 Lexicon-Based Features

Leksikon atau *lexicon* adalah kumpulan kata sentimen yang telah diketahui dan terhimpun (Desai, 2016). *Lexicon-based features* atau pendekatan berbasis leksikon dapat didefinisikan sebagai fitur berdasarkan pengetahuan yang terdapat dalam kamus. Fitur atau pendekatan ini berfokus pada pengambilan opini yang teksnya mengandung kata-kata dari leksikon kemudian diidentifikasi polaritasnya, seperti positif dan negatif. Cara kerja dari klasifikasi berbasis leksikon dijelaskan pada Gambar 2.



Gambar 2 Alur Kerja Klasifikasi Teks Berbasis Leksikon

Sumber: (Isah, 2015)

Langkah pertama yang harus dilakukan dalam proses analisis sentimen berbasis leksikon adalah dengan membuat kamus atau leksikon yang berisi kata-kata yang mengandung sentimen. Kamus ini dapat disebut dengan kamus sentimen atau *sentiment dictionary*. Terdapat beberapa pendekatan yang dapat dilakukan untuk membangun sebuah kamus sentimen, yaitu (Liu, 2012):

1. Manual

Pendekatan ini merupakan cara yang paling tidak efektif karena penyusun perlu mengumpulkan kata sentimen satu persatu sehingga menghabiskan banyak waktu. Pada akhir pengumpulan kata akan dilakukan pemeriksaan secara otomatis.

2. Berbasis kamus (*dictionary-based*)

Pendekatan ini menggunakan kata-kata dalam kamus sebagai patokan untuk mengumpulkan kata sentiment. Awalnya, kata sentiment berasal dari kata yang umum digunakan, kemudian untuk memperluas cakupan kamus sentiment, akan dicari sinonim dan antonimnya secara iterative.

3. Berbasis dokumen (*corpus based*)

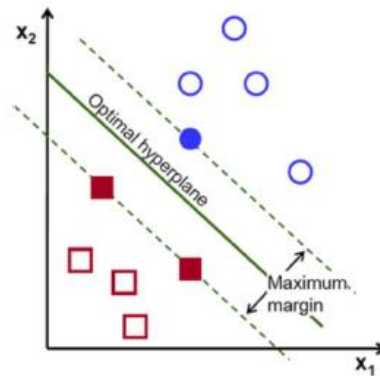
Pendekatan ini mendefinisikan kata umum yang memiliki sentimen, kemudian dicari kata sentimen lainnya berdasarkan orientasinya dari suatu domain.

Lexicon-based features atau pendekatan analisis sentimen berbasis leksikon dilakukan dengan membandingkan kata yang ada dalam suatu dokumen teks dengan kumpulan kata yang terdapat dalam leksikon. Nilai sentimen keseluruhan merupakan selisih dari

jumlah kata dengan polaritas positif dengan jumlah kata dengan polaritas negatif. Apabila skor sentimen > 0 , maka suatu dokumen teks dikatakan memiliki sentimen positif. Sebaliknya, apabila skor sentimen < 0 , maka suatu dokumen tersebut dapat dikatakan memiliki sentiment negatif. Apabila skor sentimen adalah 0 ($= 0$), maka dokumen teks dikatakan memiliki sentiment netral.

2.5 Support Vector Machine (SVM)

Support Vector Machine atau SVM adalah salah satu metode yang sering digunakan dalam memecahkan masalah *pattern recognition*, yaitu pemetaan sekumpulan data untuk mencari pola tertentu (Nugroho, Witarto, & Handoko, 2003). Konsep dasar SVM yaitu menemukan *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kategori atau kelas. Hal tersebut dapat dilakukan dengan mengukur *margin* terbesar antarkelas.



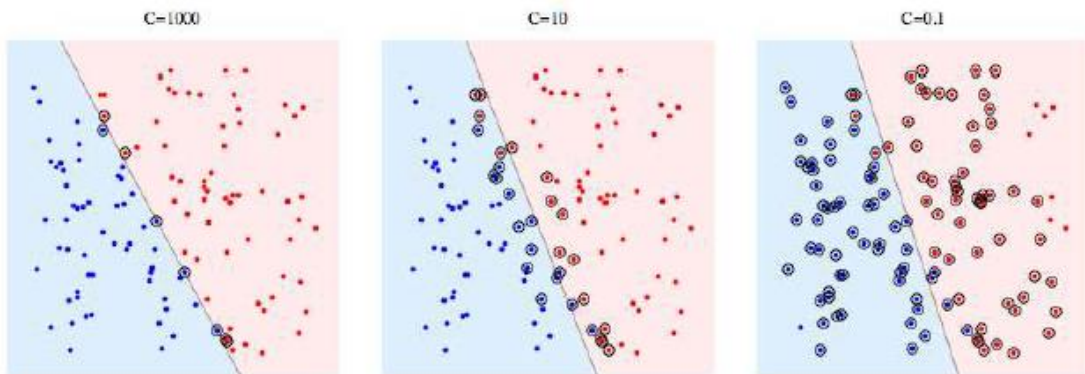
Gambar 3 Ilustrasi *Support Vector Machine* (SVM)

Sumber: (Nugroho, Witarto, & Handoko, 2003)

Gambar 3 menunjukkan bahwa terdapat beberapa observasi yang merupakan anggota dari dua kategori atau kelas, yaitu $+1$ dan -1 . SVM akan menentukan garis *hyperplane* yang terbaik untuk mengoptimalkan *margin* yang dihitung dari kedua kategori/kelas tersebut. *Margin* adalah jarak antara garis *hyperplane* dengan observasi terdekat dari masing-masing kategori atau kelas. Data yang berada pada bidang pembatas antarkategori atau antarkelas disebut sebagai *support vector*. Garis *hyperplane* terletak pada bagian tengah dari kedua kategori atau kelas yang direpresentasikan dengan *support vector*.

Salah satu parameter yang akan diuji untuk menganalisis sensitivitas model pada saat melakukan proses *text mining* adalah parameter C. Parameter C merupakan konstanta yang diatur dalam upaya regularisasi yang menjelaskan seberapa besar toleransi penalti terhadap hasil klasifikasi yang tidak tepat (Cortes & Vapnik, 1995). Parameter C adalah *trade off* dari kompleksitas aturan klasifikasi yang digunakan pada model dan frekuensi galat yang terjadi. Ilustrasi pengaruh parameter C terhadap data dokumen teks pada algoritma SVM ditunjukkan pada Gambar 4. Nilai C yang lebih kecil akan

memberi dampak terhadap model klasifikasi dengan mengorbankan separasi linear untuk mendapatkan stabilitas hasil klasifikasi.



Gambar 4 Pengaruh Parameter C terhadap Hasil Klasifikasi

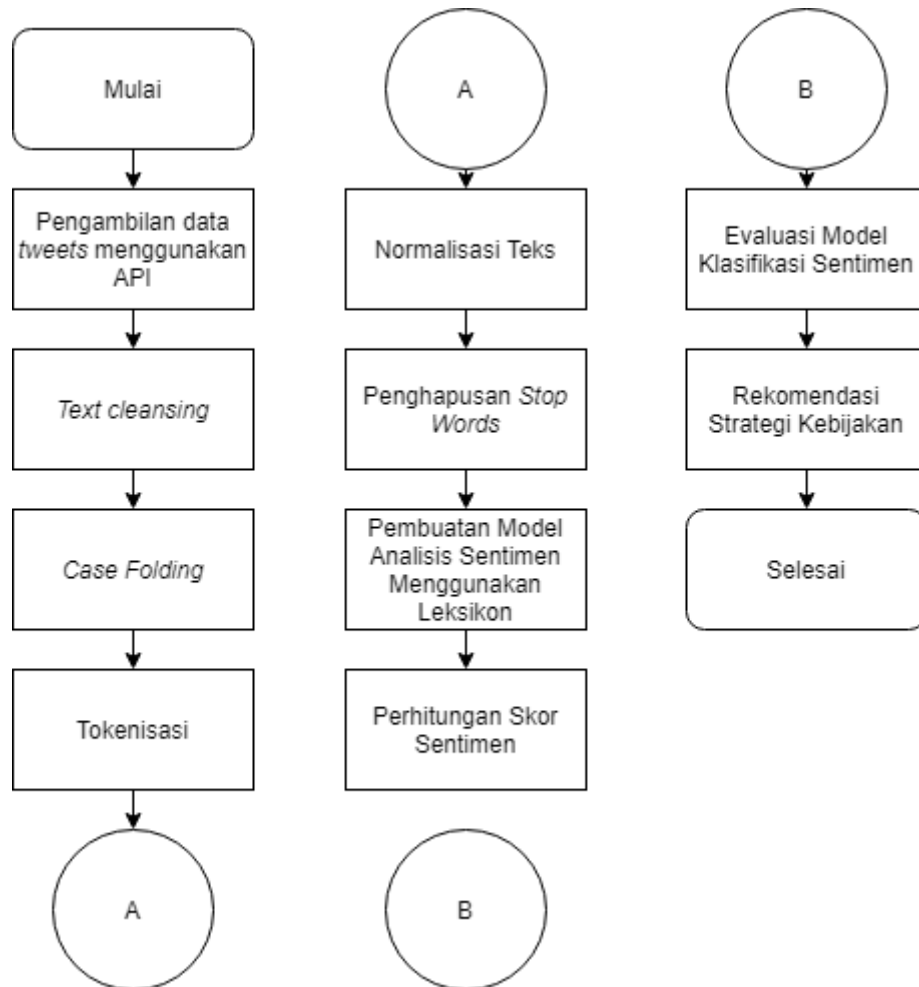
Sumber: (Bulatov, 2011)

BAB 3

METODOLOGI PENELITIAN DAN PENGOLAHAN DATA

3.1 Metodologi Penelitian

Metodologi penelitian pada kali ini ditunjukkan pada gambar berikut.



Pada penelitian analisis sentimen kali ini, data yang digunakan didapatkan dari media sosial Twitter. Dataset yang digunakan adalah data yang berisi kumpulan *tweet* yang mengandung kata kunci “Vaksin COVID Indonesia”. Dataset diambil pada tanggal 12 Desember 2020 dan berjumlah 1000 tweet terakhir yang di-post dan memiliki 16 variabel. Gambar 5 adalah contoh dataset yang diambil beserta variabelnya.

Setelah dataset didapatkan, dilakukan beberapa tahapan pengolahan data teks yang seluruhnya dilakukan pada *software* RStudio. Tahap pertama pengolahan data adalah membersihkan dan melakukan *preprocessing* pada data agar data tersebut siap diolah. Setelah diolah, dibuat visualisasi berbentuk *word cloud* untuk menunjukkan kata-kata yang paling sering keluar di dalam tweet dengan kata kunci “Vaksin COVID Indonesia”.

Kemudian, pengolahan data teks dilakukan dengan membuat fungsi untuk menghitung skor sentimen suatu *tweet* berdasarkan kamus atau leksikon. Kamus atau leksikon terbagi menjadi dua, yaitu kamus atau leksikon yang berisi kata-kata bahasa Indonesia yang bersifat positif dan kamus atau leksikon yang berisi kata-kata bahasa Indonesia yang bersifat negatif. Dari hasil pengolahan data teks, dibuat visualisasi dalam bentuk histogram yang berisi skor-skor sentimen dari *tweet* yang terdapat pada dataset. Selain itu, dibuat juga visualisasi dalam bentuk histogram yang berisi sentiment (berdasarkan skor) dari *tweet* yang terdapat pada dataset.

Tahap selanjutnya adalah dengan uji klasifikasi model yang dibuat menggunakan algoritma *Support Vector Machine (SVM) C-Classification*. Tahap ini dimulai dengan membagi dataset menjadi dua bagian, yaitu dataset *train* dan dataset *test* dengan proporsi sebanyak 75:25. Setelah itu, dilatih model yang menggunakan yang telah dibangun untuk menspesifikasikan variabel respon serta dataset *training* dan *testing*. Setelah itu, dilakukan pengujian untuk mengecek model yang telah dibuat sebelumnya. Dibuat visualisasi berbentuk *word cloud* yang berisi kata-kata pada dataset *tweet* beserta klasifikasi sentimennya (ditunjukkan dengan warna teks pada *word cloud*).

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource	screenName	retweetCount	isRetweet	retweeted	longitude	latitude
1	Alhamdulillah sudah beredar vaksin covid ya gaes! semoga s...	FALSE	0	NA	2020-12-11 12:29:12	FALSE	NA	1337373869993148416	NA	<a href="http://twitter.com/download/android" rel="nofollow...	ayusiapaaa	0	FALSE	FALSE	NA	NA
2	RT @ProfesorZubairi: Apakah kedatangan vaksin Covid-19 b...	FALSE	0	NA	2020-12-11 12:27:11	FALSE	NA	1337373362641780736	NA	<a href="http://twitter.com/download/android" rel="nofollow...	adechila	41	TRUE	FALSE	NA	NA
3	RT @sociotalker: 4 alasan kenapa vaksin covid HARUS GRAT...	FALSE	0	NA	2020-12-11 12:22:25	FALSE	NA	1337372162622017536	NA	 Twitter...	mmaulaa	124	TRUE	FALSE	NA	NA
4	RT @lawancovid19_id: Pada 6 Desember lalu, telah tiba di In...	FALSE	0	NA	2020-12-11 12:20:46	FALSE	NA	1337371747843076096	NA	<a href="http://twitter.com/download/android" rel="nofollow...	fadjroel	1	TRUE	FALSE	NA	NA
5	RT @kenno61: Ratu Elizabeth Akan Disuntik Vaksin COVID-...	FALSE	0	NA	2020-12-11 12:19:10	FALSE	NA	1337371344334307328	NA	<a href="http://twitter.com/download/android" rel="nofollow...	jitrash	10	TRUE	FALSE	NA	NA
6	RT @ProfesorZubairi: Apakah kedatangan vaksin Covid-19 b...	FALSE	0	NA	2020-12-11 12:16:23	FALSE	NA	1337370645735227392	NA	 Plum...	ji_selv	41	TRUE	FALSE	NA	NA
7	Pada 6 Desember lalu, telah tiba di Indonesia 1,2 juta dosis ...	FALSE	3	NA	2020-12-11 12:16:08	TRUE	NA	1337370580274671618	NA	<a href="http://twitter.com/download/android" rel="nofollow...	lawancovid19_id	1	FALSE	FALSE	NA	NA
8	RT @sociotalker: 4 alasan kenapa vaksin covid HARUS GRAT...	FALSE	0	NA	2020-12-11 12:11:29	FALSE	NA	1337369411280244736	NA	<a href="http://twitter.com/download/android" rel="nofollow...	lihaNov	124	TRUE	FALSE	NA	NA
9	@Haikalkurnia bagaimana prospek ekonomi dunia esp di In...	FALSE	0	Haikalkurnia	2020-12-11 12:10:53	TRUE	NA	1337369258985021441	82892265	<a href="http://twitter.com/download/iphone" rel="nofollow...	bebasdotid	0	FALSE	FALSE	NA	NA
10	RT @sociotalker: 4 alasan kenapa vaksin covid HARUS GRAT...	FALSE	0	NA	2020-12-11 12:07:40	FALSE	NA	1337368452764291073	NA	<a href="http://twitter.com/download/android" rel="nofollow...	Studio23mag	124	TRUE	FALSE	NA	NA

Gambar 5 Contoh Dataset

3.2 Data Preprocessing Teks

Setelah dataset yang berisi *tweet* didapatkan melalui Twitter dengan menggunakan API-nya, hal pertama yang harus dilakukan adalah melakukan *data preprocessing* teks. *Data preprocessing* teks dilakukan agar data teks memiliki kualitas yang baik dan seragam, sehingga proses analisis sentimen bisa dilakukan dengan optimal.

Data preprocessing teks dilakukan dengan melakukan *cleansing* (hapus *link*, hapus *retweet*, *hashtag*, dan *mention*, hapus tanda baca, emoji, dan angka, hapus *white space* dan kata kunci pencarian), *case folding*, tokenisasi, normalisasi teks (pengubahan kata tidak baku menjadi kata baku dan *stemming*), dan penghapusan *stop words*. Tabel 1 menunjukkan contoh *tweet* sebelum dan sesudah dilakukannya *data preprocessing*.

Tabel 1 Contoh *Tweet* Sebelum dan Sesudah *Data Preprocessing* Teks

Sebelum <i>Data Preprocessing</i> Teks	Sesudah <i>Data Preprocessing</i> Teks
Alhamdulillah sudah beredar vaksin covid ya gaes! semoga segera sampe indonesia https://t.co/fBwoD9erqn	alhamdulillah sudah beredar vaksin covid semoga segera sampai indonesia
Pada 6 Desember lalu, telah tiba di Indonesia 1,2 juta dosis vaksin COVID-19 dari Sinovac, sebuah kabar baik dalam... https://t.co/YdhRJVvpdU	pada desember lalu, telah tiba di Indonesia juta dosis vaksin covid dari sinovac sebuah kabar baik dalam
gratiskan vaksin Covid-19 untuk semua rakyat Indonesia! - Tandatangani Petisi! https://t.co/UkwTqmKBZd via @ChangeOrg_ID	gratis vaksin covid untuk semua rakyat indonesia tandatangani petisi via

3.2.1 Cleansing

Tahap ini merupakan tahap pembersihan data tweet dari komponen-komponen seperti simbol, tanda baca, penulisan *link* dan sebagainya yang tidak diinginkan pada data teks. Dalam melakukan *cleansing*, berikut ini adalah tahap yang dilakukan,

1. Pembersihan *link* atau URL

Data tweet yang melebihi batas karakter akan disingkat dan disimpan dalam sebuah uniform resource locator (URL). Selain itu, ada juga tweet yang memang mengandung alamat situs tertentu yang tidak dibutuhkan dalam pemrosesan data. Untuk membersihkan data teks dari URL, digunakan package ‘stringr’ dalam R dengan sintaksis yang terlampir pada Tabel 2.

2. Pembersihan *Retweet*, *Hashtag*, dan *Mention*

Retweet adalah sebuah fitur dalam Twitter untuk menyebarkan ulang *tweet* milik orang lain yang biasa ditandai dengan ‘RT’ pada awal kalimat. *Hashtag* juga adalah fitur dalam Twitter yang dapat digunakan pengguna Ketika melakukan *tweet* untuk merujuk pada suatu topik tertentu yang ditandai dengan simbol pagar

(#), sedangkan mention bermaksud untuk merujuk kepada suatu pengguna Twitter tertentu yang diawali dengan simbol (@). Untuk membersihkan data teks dari retweet, hashtags, dan mentions, digunakan package ‘stringr’ dalam R dengan sintaksis seperti pada Tabel 2.

3. Pembersihan simbol, angka, dan tanda baca

Dalam bahasa pemrograman R, “&” mengacu pada ampersand atau biasa diketahui sebagai simbol ‘&’. Tanda baca ditunjukkan dengan “[[:punct:]]” yang menunjukkan *punctuation*, dan angka ditunjukkan dengan “[[:digit:]]”. Karakter *unicode* yang digunakan untuk emoticon atau emoji ditunjukkan dengan “<U+...>” pada R. Untuk membersihkan data teks dari simbol tersebut, digunakan package ‘stringr’ dalam R dengan sintaksis seperti pada Tabel 2.

4. Pembersihan *white space*

White space adalah spasi yang tidak diperlukan yang timbul karena adanya pembersihan sebelumnya. Dilakukan juga pembersihan spasi pada awal kalimat atau akhir kalimat. Untuk membersihkan data teks dari spasi tersebut, digunakan package ‘stringr’ dalam R dengan sintaksis seperti pada Tabel 2.

Tabel 2 Sintaksis Proses *Cleansing* pada *Preprocessing* Data Teks

Proses	Sintaksis di RStudio
Pembersihan <i>link</i> atau URL	<pre>clean_tweet <- str_replace_all(unclean_tweet, "https://t.co/[a-z,A-Z,0-9]*", "") clean_tweet <- str_replace_all(clean_tweet, "http://t.co/[a-z,A-Z,0-9]*", "")</pre>
Pembersihan <i>Retweet</i> , <i>Hashtag</i> , dan <i>Mention</i>	<pre>clean_tweet <- str_replace(clean_tweet, "RT @[a-z,A-Z]*: ", "") clean_tweet <- str_replace_all(clean_tweet, "#[a-z,A-Z]*", "") clean_tweet <- str_replace_all(clean_tweet, "@[a-z,A-Z]*", "")</pre>
Pembersihan simbol, angka, dan tanda baca	<pre>clean_tweet <- gsub("&", " ", clean_tweet) clean_tweet <- gsub("[[:punct:]]", " ", clean_tweet) clean_tweet <- gsub("[[:digit:]]", " ", clean_tweet) clean_tweet <- gsub("\\<U+...>", " ", clean_tweet)</pre>
Pembersihan <i>white space</i>	<pre>clean_tweet <- stripWhitespace(clean_tweet) clean_tweet <- gsub("^ ", "", clean_tweet) # remove blank spaces at the beginning clean_tweet <- gsub(" \$", "", clean_tweet) # remove blank spaces at the end</pre>

3.2.2 Case Folding

Case folding adalah proses pengonversian seluruh data teks menjadi huruf kecil (*lower case*). Hal ini dilakukan agar seluruh data seragam sehingga tidak menimbulkan multitafsir pada program. Untuk melakukan hal ini, digunakan fungsi *tolower* dalam bahasa pemrograman R. Sintaksis yang digunakan ditunjukkan dalam Tabel 3.

Tabel 3 Sintaksis Proses *Case Folding* pada *Preprocessinng Data Teks*

Proses	Sintaksis di RStudio
<i>Case folding</i>	<code>clean_tweet <- tolower(clean_tweet)</code>

3.2.3 Tokenisasi (*Tokenization*)

Dalam tahap ini, dokumen teks berupa *tweet* yang merupakan suatu kalimat dipecah menjadi kata per kata atau token. Pemisah yang digunakan pada tahap ini adalah spasi. Sintaksis yang digunakan ditunjukkan dalam Tabel 4.

Tabel 4 Sintaksis Proses Tokenisasi pada *Preprocessinng Data Teks*

Proses	Sintaksis di RStudio
Tokenisasi	<code>word.list <- str_split(clean_tweet, '\\s+')</code> <code>words <- unlist(word.list)</code>

3.2.4 Normalisasi Teks

Tahap ini adalah tahap untuk menormalisasi data teks yang telah diolah sebelumnya. Proses normalisasi teks dilakukan dengan mengubah kata tidak baku seperti internet *slang* menjadi kata baku sesuai KBBI (PUEBI) dan menghilangkan imbuhan-imbuhan pada kata yang ada.

1. Ubah kata tidak baku menjadi tidak baku

Pada tahap ini, dilakukan pengubahan kata tidak baku menjadi tidak baku dengan memanfaatkan kamus atau leksikon yang berisi kata-kata *alay* atau kata-kata internet *slang* yang sering digunakan oleh masyarakat Indonesia terutama di media sosial. Sintaksis untuk proses ini ditunjukkan pada Tabel 5.

2. *Stemming*

Stemming adalah tahap untuk menghilangkan seluruh imbuhan yang terdapat pada suatu kata. Dengan dilakukannya proses *stemming*, maka akan dihasilkan data teks yang seragam serta memiliki makna yang sama. Hal tersebut akan mempermudah proses analisis sentiment yang menggunakan pendekatan leksikon karena leksikon yang digunakan berisi kata-kata dasar (tanpa imbuhan). Sintaksis untuk proses ini ditunjukkan pada Tabel 5.

Tabel 5 Sintaksis Proses Normalisasi Teks pada *Preprocessinng Data Teks*

Proses	Sintaksis di RStudio
Ubah kata tidak baku	<code>urlfile<-'https://raw.githubusercontent.com/nasalsabila/kamus-alay/master/colloquial-indonesian-lexicon.csv'</code>

menjadi tidak baku	<pre>lex<-read.csv(urlfile) clean_tweet <- replace_internet_slang(clean_tweet, slang = paste0("\\b", lex\$slang, "\\b"), replacement = lex\$formal, ignore.case = TRUE)</pre>
<i>Stemming</i>	<pre>clean_tweet <- as.character(clean_tweet) stemming <- function(x){ paste(lapply(x,katadasar),collapse = " ")} clean_tweet <- lapply(tokenize_words(clean_tweet[]), stemming)</pre>

3.2.5 Penghapusan *Stop Words*

Stop words yang digunakan berasal dari penelitian sebelumnya (Tala, 2003). Daftar kata *stop words* yang digunakan pada proses *data preprocessing* teks berjumlah 758 kata. Contoh *stop words* yang digunakan terlampir pada Tabel 6. Sintaksis untuk melakukan penghapusan *stop words* ditunjukkan pada Tabel 7.

Tabel 6 Contoh *Stop Words* yang Digunakan

No	<i>Stop words</i>
1	ada
2	adalah
3	adanya
4	adapun
5	agak
6	agakny
7	agar
8	akan
9	akankah
10	akhir

Tabel 7 Sintaksis Proses Penghapusan *Stop Words* pada *Preprocessing Data Teks*

Proses	Sintaksis di RStudio
Penghapusan <i>stop words</i>	<pre>urlfile<-"https://raw.githubusercontent.com/masdevid/ID- Stopwords/master/id.stopwords.02.01.2016.txt" stopwordsbahasa<-read.delim(urlfile, header = FALSE, stringsAsFactors=FALSE) colnames(stopwordsbahasa) <- "stopwordsbahasaid" stopwordsbahasa <- as.character(stopwordsbahasa\$stopwordsbahasaid)</pre>

	<code>clean_tweet <- removeWords(clean_tweet, stopwordsbahasa)</code>
--	--

3.3 Analisis Sentimen

Pada penelitian ini, analisis sentimen dilakukan dengan menggunakan pendekatan berbasis leksikon atau kamus. Secara umum, analisis sentimen dapat didefinisikan sebagai proses klasifikasi terhadap suatu dokumen teks berdasarkan sentimen yang terkandung pada teks tersebut. Dalam konteks penelitian ini, dokumen teks yang dimaksud adalah *tweet*. Untuk lebih mengetahui konteks kata-kata yang akan diteliti pada analisis sentimen, dibuatlah *word cloud* atas data teks yang telah mendapat perlakuan *data preprocessing* teks. Gambar 6 adalah *word cloud* dari data teks yang telah diolah. Sintaksis untuk membuat *word cloud* ditunjukkan pada Tabel 8.



Gambar 6 *Word Cloud* dari Dokumen *Tweet* terkait Vaksin COVID Indonesia

Tabel 8 Sintaksis Pembuatan *Word Cloud* pada Data yang Telah Diolah

Proses	Sintaksis di RStudio
Pembuatan <i>word cloud</i> dari data yang	<pre>lex.corpus <- Corpus(VectorSource(clean_tweet)) lex.dtm <- DocumentTermMatrix(lex.corpus) m <- as.matrix(lex.dtm) v <- sort(colSums(m), decreasing=TRUE)</pre>

telah diolah	<pre> d <- data.frame(word = names(v), freq=v) wordcloud(words= d\$word, freq=d\$freq, min.freq = 1, max.words = 100, random.order=FALSE, rot.per=0.35, colors = brewer.pal (name = "Dark2", 50), scale=c(2.5,0.5)) </pre>
--------------	---

Setelah itu, dilakukan proses pembuatan model skor sentimen. Terdapat tiga klasifikasi yang digunakan pada penelitian ini dengan berlandaskan skor suatu *tweet*, yaitu sentimen positif (skor sentimen > 0), negatif (skor sentimen < 0), dan netral (skor sentimen = 0). Sintaksis yang digunakan untuk membuat model skor sentimen adalah sebagai berikut.

Tabel 9 Sintaksis Proses Pembuatan Model Skor Sentimen

Proses	Sintaksis di RStudio
Pembuatan model skor sentimen	<pre> # Buka file terkait sentimen pos.words <- scan("D:/Apiq/Tingkat 2, 3, 4 MRI/Semester 7/Data Science and Machine Learning/positive-words.csv", what = 'character') neg.words <- scan("D:/Apiq/Tingkat 2, 3, 4 MRI/Semester 7/Data Science and Machine Learning/negative-words.csv", what = 'character') # Buat fungsi untuk analisis sentiment sentiment.score = function(sentences, pos.words, neg.words, .progress = 'none') { # Akan dihasilkan vektor yang berisi kalimat. Plyr akan mengurus list # atau vektor sebagai "l" # dibutuhkan array sederhana ("a") yang berisi skor, maka digunakan # "l" + "a" + "ply" = "laply": scores = laply(sentences, function(sentence, pos.words, neg.words) { </pre>

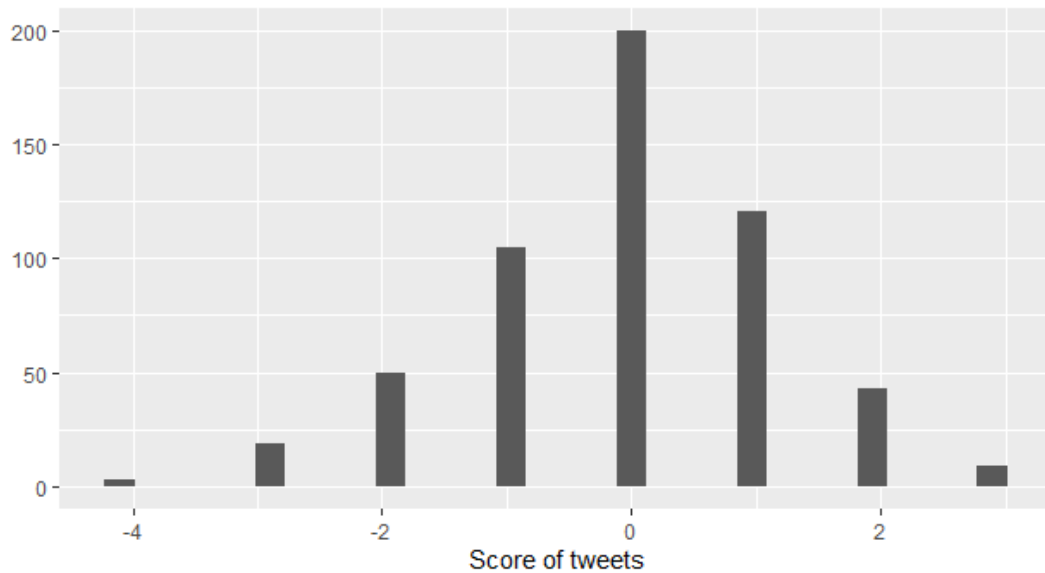
	<pre> # bersihkan kalimat dengan R's regex-driven global substitute, gsub(): sentence = gsub('[:punct:]]', '', sentence) sentence = gsub('[:cntrl:]]', '', sentence) sentence = gsub('\\d+', '', sentence) # ubah seluruh huruf menjadi huruf kecil: sentence = tolower(sentence) # pisahkan kalimat menjadi kata word.list = str_split(sentence, '\\s+') # terkadang list memiliki hierarki yang lebih tinggi sebanyak satu level words = unlist(word.list) # bandingkan kata-kata dengan kamus atau leksikon positif dan negatif pos.matches = match(words, pos.words) neg.matches = match(words, neg.words) # match() menghasilkan posisi dari kata-kata yang dicek atau NA # hasilnya adalah TRUE/FALSE: pos.matches = !is.na(pos.matches) neg.matches = !is.na(neg.matches) # TRUE/FALSE akan dihitung 1/0 oleh sum(): score = sum(pos.matches) - sum(neg.matches) return(score) }, pos.words, neg.words, .progress=.progress) scores.df = data.frame(score=scores, text=sentences) return(scores.df) } # Aplikasikan fungsi sentimen ke dalam tweets result <- sentiment.score(unclean_tweet,pos.words,neg.words) </pre>
--	--

Dengan dibuatnya model skor sentimen, maka dapat dilakukan perhitungan skor sentimen dan jumlah sentimen terhadap seluruh dataset teks yang ada. Contoh

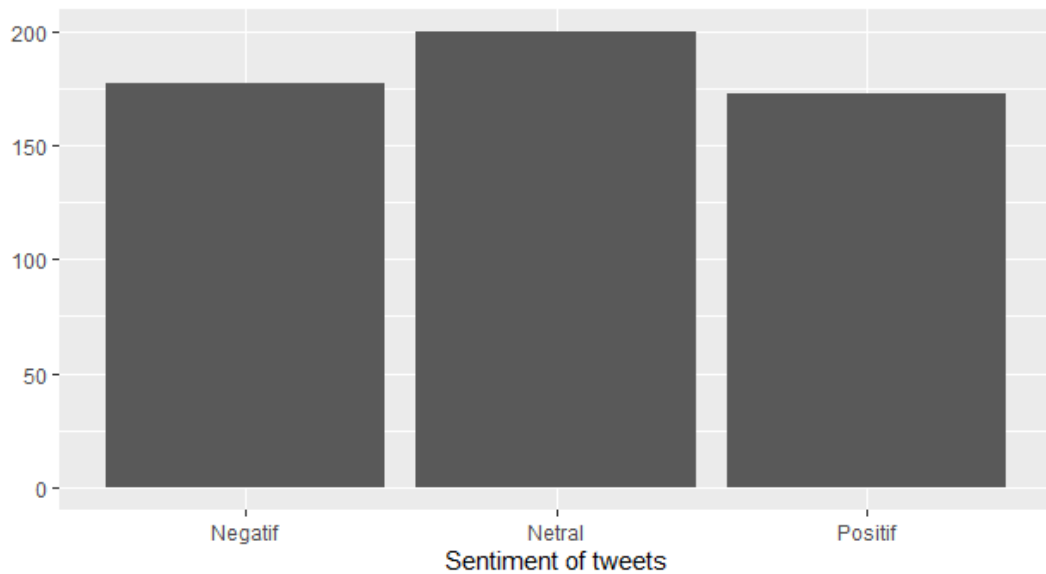
klasifikasi sentimen ditunjukkan pada Tabel 10. Visualisasi dari jumlah skor terkait data teks *tweet* ditunjukkan pada Gambar 7 dan visualisasi dari jumlah sentiment terkait data teks *tweet* ditunjukkan pada Gambar 8. Sintaksis yang digunakan untuk menampilkan kedua visualisasi tersebut ditunjukkan pada Tabel 11.

Tabel 10 Contoh Klasifikasi Sentimen Data Teks

<i>Tweet</i>	Klasifikasi Sentimen
alhamdulillah sudah beredar vaksin covid ya! semoga segera sampe indonesia	Positif
apakah datang vaksin covid bisa jadi game changer bisa saja tapi tidak dalam jangka pendek herd immunity itu	Netral
reaksi alergi vaksin covid inggris saran orang dengan riwayat alergi serius hindari vaksin pfizer	Negatif



Gambar 7 Skor Sentimen Data Teks (*Tweets*)



Gambar 8 Jumlah Sentimen terkait Data Teks (*Tweets*)

Tabel 11 Sintaksis Pembuatan Visualisasi Histogram Skor Sentimen dan Jumlah Sentimen Data Teks (*Tweets*)

Proses	Sintaksis di RStudio
Visualisasi histogram skor sentiment data teks (<i>tweets</i>) dan jumlah sentiment terkait data teks (<i>tweets</i>)	<pre># Visualisasi #score hist(result\$score, col = "yellow", main = 'Score of tweets', ylab = 'Count of tweets') count(result\$score) qplot(result\$score,xlab = "Score of tweets") #sentiment hist(table.sentiment\$sentimen, col = "yellow", main = 'Sentiment of tweets', ylab = 'Count of tweets') count(table.sentiment\$sentimen) qplot(table.sentiment\$sentimen,xlab = "Sentiment of tweets")</pre>

3.4 Evaluasi Analisis Sentimen

Evaluasi analisis sentimen dilakukan dengan melakukan uji dari hasil analisis sentimen dengan menggunakan data *test* yang telah dipisahkan sebelumnya dengan proporsi 25:75 bila dibandingkan dengan data *train*. Hasil klasifikasi sentimen oleh model analisis sentimen untuk data *test* akan dibandingkan dengan hasil dari data *train* untuk memeriksa akurasi model yang telah dibuat.

Hal yang pertama dilakukan pada proses uji analisis sentiment adalah dengan membuat *document term matrix* (DTM). *Document term matrix* adalah matriks yang menunjukkan setiap kata yang terdapat pada *tweet* dengan jumlah dari kata tersebut pada setiap dokumen teks. Uji hasil klasifikasi dilakukan dengan menggunakan algoritma SVM. Algoritma ini adalah algoritma yang lazim dipakai untuk proses klasifikasi teks dan memiliki performa yang baik. Uji model klasifikasi sentimen dilakukan dengan menggunakan *package* RTextTools. Sintaksis yang digunakan untuk proses uji klasifikasi ditunjukkan pada Tabel 12.

Tabel 12 Sintaksis Uji Model Klasifikasi Sentimen

Proses	Sintaksis di RStudio
Uji model klasifikasi sentimen	<pre> # Buat matriks DTM matrix= create_matrix(table.sentiment[,1]) mat = as.matrix(matrix) # Latih model # SVM # Bangun data untuk menspesifikasi variabel respon, dataset training dan testing container = create_container(matrix, as.numeric(as.factor(table.sentiment[,1])), inTrain,virgin=FALSE) models = train_models(container, algorithms="SVM") results = classify_models(container, models) # Recall accuracy recall_accuracy(as.numeric(as.factor(table.sentiment\$sentimen[length(table.sentiment\$sentimen)- inTrain])), results["SVM_LABEL"]) # Rangkuman model analytics = create_analytics(container, results) analytics@algorithm_summary </pre>

Uji performansi lainnya dilakukan dengan menghitung akurasi dari model klasifikasi sentimen yang dibuat. Metode yang digunakan untuk mengukur akurasi adalah metode SVM *C-Classification* yang lazim digunakan untuk model klasifikasi dengan nilai 1 pada parameter C dan kernel linear. Akurasi didapat dengan menggunakan *k-fold cross-validation* dengan $k = 9$. Rata-rata akurasi yang diperoleh pada uji performansi akurasi adalah 96,57%. Selain itu, dihitung juga nilai ukuran performa lainnya

Proses	Sintaksis di RStudio
Pembuatan <i>word cloud</i> beserta klasifikasi sentimennya	<pre> # Buat wordcloud yang berisi kata-kata beserta klasifikasi sentimennya lex.corpus <- Corpus(VectorSource(table.sentiment\$text)) lex.dtm <- DocumentTermMatrix(lex.corpus) m <- as.matrix(lex.dtm) v <- sort(colSums(m),decreasing=TRUE) d <- data.frame(word = names(v),freq=v) for(i in 1:nrow(d)){ word.sentimen <- c() index.sentimen <- c() tweet.sentimen <- table.sentiment\$sentimen[grep(d\$word[i], table.sentiment\$text)] positif <- sum(tweet.sentimen=="Positif") negatif <- sum(tweet.sentimen=="Negatif") netral <- sum(tweet.sentimen=="Netral") word.sentimen <- rbind(positif,negatif,netral) index.sentimen <- which(word.sentimen==max(positif,negatif,netral)) if (index.sentimen == 1){ d\$kelas[i] <- "Positif" } else if (index.sentimen == 2){ d\$kelas[i] <- "Negatif" } else { d\$kelas[i] <- "Netral" } } wc <- wordcloud(words = d\$word, freq = d\$freq, min.freq = 1, max.words=100, random.order=FALSE, rot.per=0.35, ordered.colors=TRUE, colors=c("#FF0033", "#FFCC33", "#99CCFF")[factor(d\$kelas)], scale=c(2.5,0.5)) </pre>

BAB 4

ANALISIS

4.1 Analisis terkait Sentimen Rakyat Melalui *Tweet* tentang Vaksin COVID-19 di Indonesia

Skor sentimen dari pendapat rakyat Indonesia yang didapat melalui Twitter terkait vaksin COVID-19 di Indonesia didapat dengan membuat model analisis sentiment dengan menggunakan pendekatan leksikon (*lexicon-based features*). Total pendapat rakyat (dalam bentuk *tweet*) yang didapat adalah 1000 *tweets* (dokumen teks).

Dari hasil pengolahan data, didapat bahwa skor sentimen dari pendapat rakyat (melalui *tweet*) terkait COVID-19 di Indonesia memiliki rentang dari -4 hingga +3. Berdasarkan Gambar 7, tiga skor terbanyak yang didapat dari data teks yang telah diolah adalah 0, +1, dan -1. Secara keseluruhan, rangkuman terkait skor sentimen data teks yang telah diolah ditunjukkan pada Gambar 10.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-4.00000	-1.00000	0.00000	-0.07273	1.00000	3.00000

Gambar 10 Rangkuman Skor Sentimen Data Teks yang Telah Diolah

Berdasarkan rangkuman terkait skor sentimen tersebut, maka dapat dikatakan bahwa sentimen rakyat Indonesia (disampaikan dalam bentuk *tweet*) terkait COVID-19 di Indonesia adalah negatif karena memiliki rata-rata skor sentimen -0.07273 (terlampir pada Gambar 10).

Dengan diketahuinya hal tersebut, penulis melakukan verifikasi ulang terhadap *tweets* di Twitter yang berkaitan dengan vaksin COVID-19 di Indonesia. Data yang didapat tersebut ternyata sesuai dengan kenyataan di Twitter karena memang secara umum *tweet* bersifat organik (dilakukan oleh manusia asli) yang berkaitan dengan kata kunci “Vaksin COVID-19 Indonesia” berisi keluhan atau kritik terhadap kebijakan pemerintah terkait vaksin COVID-19.

Pada umumnya, keluhan dan kritik pada *tweet* terkait COVID-19 di Indonesia berisi tentang uji klinis vaksin yang masih belum jelas dan biaya vaksin yang mahal atau tidak terjangkau oleh rakyat kecil. Terkait uji klinis, Indonesia memang dapat dikatakan terburu-buru melakukan pengadaan vaksin karena hingga saat ini belum ada vaksin yang melewati uji klinis tingkat 3 (batas normal untuk disebarkannya suatu vaksin secara masif). Sehingga, vaksin yang dibeli dan disediakan oleh pemerintah dapat dikatakan masih prematur karena belum melewati uji klinis tingkat 3. Dengan kata lain, vaksin tersebut belum tentu memiliki performa yang baik untuk membunuh virus COVID-19.

Terkait biaya, kritik dan keluhan rakyat Indonesia melalui Twitter pada umumnya berkaitan dengan tingginya harga vaksin, terutama dari sudut pandang rakyat kecil. Hal tersebut akan mempersulit proses penggunaan vaksin oleh rakyat Indonesia karena tidak seluruh golongan dapat mengaksesnya. Padahal ide utama vaksin adalah tersebarinya orang-orang yang menggunakan vaksin itu sendiri, sehingga terbentuk imunitas yang masif dari suatu komunitas (*herd immunity*).

Hal tersebut akan bertentangan tujuan awal dilakukannya pengadaan vaksin oleh pemerintah Indonesia untuk mendukung kondisi ekonomi yang sempat terpuruk karena bisa jadi vaksin tersebut memiliki efek samping yang berbahaya bagi orang yang menggunakannya. Efek samping tersebut justru dapat membuat seseorang menjadi tidak produktif dan mengimplikasikan tidak adanya perkembangan terkait kondisi ekonomi di Indonesia.

Selain itu, tingkat keefektifan vaksin juga belum dapat dipastikan kualitasnya. Dengan kata lain, seseorang belum tentu imun terhadap COVID-19 setelah satu kali melakukan vaksin. Hal tersebut mengharuskan seseorang untuk menerima vaksin kemungkinan besar lebih dari satu kali. Padahal biaya yang dibutuhkan untuk mendapatkan vaksin tergolong cukup tinggi, apalagi jika dari sudut pandang rakyat kecil. Oleh karena itu, pendapatan mereka akan semakin tergerus untuk konsumsi hal yang sebenarnya merupakan kebutuhan dasar mereka dari pemerintah di masa pandemi COVID-19, yaitu vaksin.

Berdasarkan hal-hal tersebut, maka dapat dikatakan bahwa strategi terkait kebijakan vaksin COVID-19 di Indonesia adalah terkait pengadaan vaksin COVID-19 itu sendiri dan pendistribusian vaksin tersebut. Pengadaan vaksin COVID-19 harus diseleksi secara ketat dan harus memenuhi kriteria bahwa vaksin sudah melewati uji klinis tingkat 3, sehingga dapat dijamin keefektifannya dan keamanannya. Pendistribusian vaksin COVID-19 harus merata baik secara wilayah maupun orang yang menerimanya. Hal tersebut dapat dicapai dengan menggratiskan biaya vaksin COVID-19, sehingga seluruh rakyat Indonesia dapat merasakan manfaat adanya vaksin COVID-19. Pendistribusian vaksin COVID-19 yang tidak merata akan sangat tidak efektif dalam menciptakan lingkungan yang imun terhadap COVID-19 karena seperti yang telah disebutkan sebelumnya, ide utama sebuah vaksin adalah tercapainya *herd immunity* di mana setiap orang imun terhadap suatu virus.

4.2 Analisis Rekomendasi Strategi Kebijakan Vaksin COVID-19 di Indonesia

Pada subbab sebelumnya, telah disebutkan rekomendasi kebijakan terkait vaksin COVID-19 di Indonesia yaitu dengan memperketat seleksi masuknya vaksin (harus melewati uji klinis tingkat 3) dan pendistribusian vaksin yang merata (terutama dari segi biaya yang harusnya gratis). Berdasarkan rekomendasi kebijakan tersebut, penulis

memberikan rekomendasi strategi untuk pengimplementasiannya. Rekomendasi yang diberikan adalah sebagai berikut.

1. Rekomendasi Seleksi Vaksin

Pemerintah melalui Kementerian Kesehatan harus membuat peraturan atau payung hukum yang mewajibkan seluruh entitas (termasuk pemerintah itu sendiri) yang ingin melakukan pengadaan vaksin untuk memenuhi peraturan atau payung hukum tersebut. Peraturan atau payung hukum tersebut berisi *guideline* terkait vaksin yang boleh beredar di Indonesia. Secara rinci harus dinyatakan kriteria-kriteria terkait vaksin yang diterima peredarannya di Indonesia.

Hal utama yang harus dipenuhi terkait seleksi vaksin adalah tentang tingkat uji klinis. Suatu vaksin yang masuk ke Indonesia harus melewati uji klinis tingkat 3 agar terjamin keefektifan dan keamanannya. Jika vaksin telah melewati uji klinis tingkat 3, maka dapat dipastikan vaksin akan berdampak terhadap kehidupan seseorang secara signifikan, baik secara kesehatan maupun secara ekonomi.

2. Rekomendasi Pendistribusian Vaksin

Pendistribusian suatu vaksin di suatu negara akan berkaitan erat dengan harga serta distribusi itu sendiri. Oleh karena itu, rekomendasi strategi akan diberikan terkait kedua hal tersebut.

Terkait harga vaksin, jika vaksin berasal dari entitas swasta, maka harus dibuat suatu batas harga agar tidak terjadi pengambilan keuntungan yang sangat tinggi dari komoditas yang tergolong kebutuhan dasar di masa pandemi COVID-19. Jika vaksin berasal dari pemerintah, maka vaksin harusnya digratiskan karena hal tersebut dapat menjadi stimulus untuk bergerak ekonomi secara normal kembali. Indonesia dapat mencontoh Kanada yang membuat kebijakan bahwa setiap warga negaranya berhak mendapatkan vaksin COVID-19 secara gratis. Dengan begitu, dapat dicapai *herd immunity* secara efektif.

Terkait pendistribusian vaksin, jika vaksin berasal dari entitas swasta, maka pemerintah dapat mengatur area distribusinya. Sehingga tidak ada perang harga antar satu distributor vaksin dengan distributor yang lainnya. Hal tersebut dapat membuat vaksin diakses lebih mudah oleh rakyat Indonesia. Jika pendistribusian vaksin yang dilakukan oleh pemerintah, seharusnya dapat dilakukan secara menyeluruh ke penjuru-penjuru kecil negara Indonesia. Hal tersebut dapat dicapai dengan memanfaatkan sarana dan prasarana puskesmas serta tenaga medis yang terdapat di dalamnya.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Skor sentimen *tweet* terkait vaksin COVID-19 di Indonesia memiliki rentang dari -4 hingga +3 dengan rata-rata -0,07273 yang menandakan sentimen rakyat Indonesia terkait COVID-19 di Indonesia adalah negatif.
2. Klasifikasi sentimen yang jumlahnya paling banyak adalah netral.
3. Klasifikasi sentimen yang jumlahnya paling sedikit adalah positif.
4. Rekomendasi terkait strategi yang berkaitan dengan kebijakan vaksin COVID-19 di Indonesia terlampir pada Subbab 4.2. Analisis Rekomendasi Strategi Kebijakan Vaksin COVID-19 di Indonesia.

5.2 Saran Penelitian

Penelitian analisis sentimen ini dilakukan hanya dengan menggunakan pendekatan leksikon, sehingga akan lebih baik jika ada penelitian serupa yang menggunakan metode SVM, Naïve Bayes, dan sebagainya. Dengan begitu, dapat diketahui model yang paling optimal dalam mengklasifikasikan sentimen dari *tweets* yang berisi pendapat rakyat Indonesia terkait COVID-19. Selain itu, saran penelitian lainnya adalah dengan memberikan rekomendasi strategi kebijakan yang holistik, baik dari segi kesehatan hingga segi operasional pendistribusian vaksinnya.

DAFTAR PUSTAKA

- Bulatov, Y. (2011, Januari 8). *SVM - hard or soft margins?* Retrieved from Github: <http://yaroslavvb.com/upload/save/so-libsvm.png>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 273-297.
- Desai, M. &. (2016). Techniques for Sentiment Analysis of Twitter. *International Conference on Computing*.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 77-82.
- Feinerer, I. (2008). *A text mining framework in R and its applications*. Vienna: WU Vienna University of Economics and Business.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced*. New York: Cambridge University Press.
- Hotho, A., Nurnberger, A., & Neagu, D. (2005). *A Brief Survey of Text Mining*. Kassel: University of Kassel.
- Isah, H. T. (2015). Social Media Analysis for Product Safety. *Computational Intelligence (UJKCI), 2014 14th UK Workshop on IEEE*, 1-7.
- Kompas. (2020, December 13). *Daftar dan Rincian Perkiraan Harga Vaksin Covid-19 di Indonesia*. Retrieved from Kompas: <https://www.kompas.tv/article/130420/daftar-dan-rincian-perkiraan-harga-vaksin-covid-19-di-indonesia>
- Larose, D. (2005). In *Discovering Knowledge in Data: An Introduction to Data*. Hoboken: John Wiley & Sons, Inc.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Nugroho, A., Witarto, A., & Handoko, D. (2003). Application of Support Vector. *Indonesian Scientific Meeting in Central Japan*.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 1-135.
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. . Amsterdam: Institute for Logic, Language and Computation. Universiteit van Amsterdam, The Netherlands.