

***SENTIMENT ANALYSIS PERSEPSI MASYARAKAT TERHADAP  
PROTOKOL KESEHATAN VIRUS CORONA DI TWITTER  
MENGUNAKAN WORD2VEC MODEL DAN RECURRENT NEURAL  
NETWORK LEARNING***

**(SKRIPSI)**

**Oleh**

**NI PUTU AYU ANESCA  
NPM 1617051126**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

***SENTIMENT ANALYSIS PERSEPSI MASYARAKAT TERHADAP  
PROTOKOL KESEHATAN VIRUS CORONA DI TWITTER  
MENGUNAKAN WORD2VEC MODEL DAN RECURRENT NEURAL  
NETWORK LEARNING***

**(SKRIPSI)**

**Oleh**

**NI PUTU AYU ANESCA  
NPM 1617051126**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

## **ABSTRACT**

### **SENTIMENT ANALYSIS OF PUBLIC PERCEPTION OF THE CORONA VIRUS HEALTH PROTOCOL ON TWITTER USING WORD2VEC MODEL AND RECURRENT NEURAL NETWORK LEARNING**

**By**

**Ni Putu Ayu Anesca**

*Sentiment analysis is a computational study of opinion from various opinions, which is part of the work that conducts a review related to the computational treatment of opinions, sentiments, and perceptions of the text. To solve various problems in sentiment analysis, needed a good text representation method. In this study, a deep learning analysis was carried out using the Recurrent Neural Network (RNN) method and the Word2Vec Model as word embedding in sentiment classification. The sentiment dataset used comes from user reviews on Twitter (tweets) on the health protocols implemented by the public from the government's appeal. The results showed that the RNN model using sigmoid activation resulted in the greatest accuracy of 66%. The training process in this test uses 10 epochs and 32 batch sizes so that the precision value for negative sentiment is 54% and for positive sentiment is 67%.*

*Keywords: Health Protocol, RNN, Sentiment Analysis, Word2Vec Model*

## ABSTRAK

# ***SENTIMENT ANALYSIS PROTOKOL KESEHATAN VIRUS CORONA DARI TWEET MENGGUNAKAN WORD2VEC MODEL DAN RECURRENT NEURAL NETWORK LEARNING***

Oleh

Ni Putu Ayu Anesca

*Sentiment analysis* adalah studi komputasi opini dari berbagai opini, yaitu bagian dari pekerjaan yang melakukan peninjauan berkaitan dengan perlakuan komputasional opini, sentimen, dan persepsi dari teks. Untuk menyelesaikan berbagai masalah dalam *sentiment analysis*, diperlukan metode representasi teks yang baik. Dalam penelitian ini, dilakukan analisis *deep learning* menggunakan metode *Recurrent Neural Network* (RNN) serta *Word2Vec Model* sebagai *embedding* kata dalam klasifikasi *sentiment*. Dataset *sentiment* yang digunakan berasal dari *tweet* pengguna di Twitter (*tweet*) terhadap protokol kesehatan yang diterapkan oleh masyarakat dari himbauan pemerintah Indonesia. Hasil penelitian menunjukkan bahwa Model RNN menggunakan aktivasi *sigmoid* menghasilkan akurasi terbesar bernilai 66%. Proses *training* pada pengujian ini menggunakan sebanyak 10 *epoch* dan 32 *batch size* sehingga nilai *precision* untuk *sentiment* negatif sebesar 54% dan untuk *sentiment* positif sebesar 67%.

Kata kunci: Protokol Kesehatan, RNN, *Sentiment Analysis*, *Word2Vec Model*

***SENTIMENT ANALYSIS PERSEPSI MASYARAKAT TERHADAP  
PROTOKOL KESEHATAN VIRUS CORONA DI TWITTER  
MENGUNAKAN WORD2VEC MODEL DAN RECURRENT NEURAL  
NETWORK LEARNING***

**Oleh**

**NI PUTU AYU ANESCA**

**Skripsi**

**Sebagai Salah Satu Syarat Untuk Mencapai Gelar  
SARJANA KOMPUTER**

**Pada**

**Jurusan Ilmu Komputer  
Fakultas Matematika dan Ilmu Pengetahuan Alam**

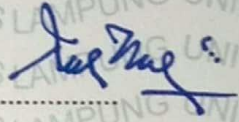


**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

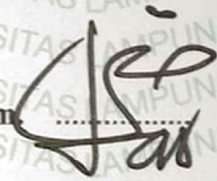
**MENGESAHKAN**

**1. Tim Penguji**

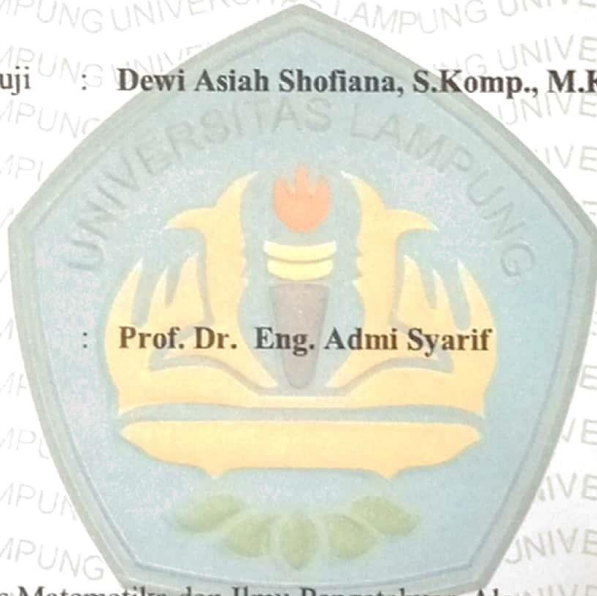
**Ketua Penguji : Dr. Ir. Kurnia Muludi, M.S.Sc.**



**Penguji I  
Sekretaris Penguji : Dewi Asiah Shofiana, S.Komp., M.Kom**



**Penguji II  
Penguji Utama : Prof. Dr. Eng. Admi Syarif**



**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



**Dr. Eng. Suropto Dwi Yuwono, M.T.**  
NIP 19740705 200003 1 001



**Tanggal Lulus Ujian Skripsi : 9 Februari 2022**

Judul Skripsi : **SENTIMENT ANALYSIS PERSEPSI MASYARAKAT TERHADAP PROTOKOL KESEHATAN VIRUS CORONA DI TWITTER MENGGUNAKAN WORD2VEC MODEL DAN RECURRENT NEURAL NETWORK LEARNING**

Nama Mahasiswa : **Ni Putu Ayu Anesca**

Nomor Pokok Mahasiswa : **1617051126**

Program Studi : **SI Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**


**MENYETUJUI,**

1. **Komisi Pembimbing**

  
Dr. Ir. Kurnia Muludi, M.S.Sc.  
NIP 19640616 198902 1 001

  
Dewi Asiah Shofiana, S.Komp., M.Kom.  
NIP 19950929 202012 2 030

2. **Ketua Jurusan Ilmu Komputer**

  
Didik Kurniawan, M.T.  
NIP 19800419 200501 1 004

## PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa skripsi saya yang berjudul “*Sentiment Analysis* Persepsi Masyarakat Terhadap Protokol Kesehatan Virus Corona di Twitter Menggunakan *Word2vec Model* dan *Recurrent Neural Network Learning*” merupakan hasil karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang saya terima.

Bandar Lampung, 9 Februari 2022  
Yang Menyatakan,



**Ni Putu Ayu Anesca**  
NPM. 1617051126



## PERSEMBAHAN

Puji syukur atas kehadiran Tuhan Yang Maha Esa atas segala berkah-Nya, sehingga penyusunan skripsi ini dapat terselesaikan.

Saya persembahkan karya ini kepada:

*Ayahanda, Ibunda dan Adikku Tersayang*

Ibu dan ayah yang luar biasa saya cintai dan saya sayangi. Orang yang telah membesarkan, mendidik, menjaga, melindungi, dan memberikan dukungan untuk anak-anaknya. Orang yang selalu memberikan doa agar saya dalam melakukan setiap langkah diberikan kemudahan dan kesuksesan untuk nantinya berguna bagi saya sendiri dan khayalak banyak.

Kepada keluarga besar dan teman-teman yang selalu mendukung dan mendoakan saya.

Serta Almamater Tercinta Universitas Lampung.

## **MOTTO**

**“Semoga semua pikiran yang baik datang dari semua penjuru.”**

(Reg Weda I.89.1)

**“When things get hard, stop for a while and look back, see how far you’ve come. Don’t forget how regarding it is.”**

(Kim Taehyung)

**“Be yourself, that’s the best for you.”**

(Ni Putu Ayu Anesca)

## SANWACANA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan izin dan kekuatan, sehingga penulis dapat menyelesaikan skripsi dengan judul “***Sentiment Analysis Persepsi Masyarakat Terhadap Protokol Kesehatan Virus Corona di Twitter Menggunakan Word2Vec Model dan Recurrent Neural Network Learning***” sebagai salah satu syarat untuk memperoleh gelar sarjana pada Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Penyusunan skripsi ini dapat terselesaikan dengan baik berdasarkan bantuan, bimbingan dan saran dari berbagai pihak. Pada kesempatan kali ini penulis ingin menyampaikan terima kasih kepada:

1. Sang Hyang Widhi Wasa yang telah memberikan rahmat dan karunia-Nya sehingga dapat menyusun penulisan skripsi ini dengan baik.
2. Papa, Mama dan Adik tercinta di rumah yang selalu mendoakan, mendukung dan memberi semangat.
3. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. selaku Dosen Pembimbing Pertama atas bimbingan, nasihat, ide dan motivasi yang luar biasa dalam membantu penyusunan skripsi ini hingga diselesaikan.
4. Ibu Dewi Asiah Shofiana, S.Komp., M. Kom. selaku Dosen Pembimbing Kedua atas bimbingan, saran, dan semangat dalam membantu penyusunan skripsi ini hingga diselesaikan.
5. Bapak Prof. Dr. Eng. Admi Syarif selaku Pembahas atas bimbingan, kritik dan saran dalam membantu penyusunan skripsi ini hingga diselesaikan.
6. Bapak Bambang Hermanto, S.Kom., M.Cs., selaku Pembimbing Akademik penulis atas nasihat, motivasi, dan arahan kepada penulis hingga penulis dapat menempuh pendidikan dengan baik di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Dr. Eng. Suripto Dwi Yuwono selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
9. Ibu Astria Hijriani, S.Kom., M.Kom., selaku Sekretaris Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
10. Seluruh Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan ilmu dan pelajaran hidup selama penulis menjadi mahasiswa.
11. Anggie Angraeni, Ajijah, dan Friska Daesy Elvina Simbolon sebagai teman terbaik yang memberikan kritik, saran dan dukungan selama menempuh pendidikan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
12. Hanif Kurniawan yang memberikan semangat dan motivasi selama menempuh pendidikan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
13. Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hoseok, Park Jimin, Kim Taehyung dan Jeon Jungkook atas hiburan dan motivasi sebagai inspirasi melalui karya-karyanya saat penulisan skripsi hingga diselesaikan.
14. Semua pihak yang telah membantu penulis yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa hasil penulisan skripsi ini masih banyak kekurangan dan kelemahan, baik dalam isi maupun sistematikanya. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran yang membangun untuk menyempurnakan. Semoga skripsi ini dapat memberikan manfaat bagi penulis, pembaca dan semua pihak.

Bandar Lampung, 9 Februari 2022  
Penulis,

**Ni Putu Ayu Anesca**  
NPM. 1617051126

## DAFTAR ISI

Halaman

<b>DAFTAR ISI</b> .....	<b>xiii</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>DAFTAR GAMBAR</b> .....	<b>xvi</b>
<b>I. PENDAHULUAN</b> .....	<b>1</b>
A. Latar Belakang.....	1
B. Rumusan Masalah.....	3
C. Batasan Masalah.....	3
D. Tujuan Penelitian.....	3
E. Manfaat Penelitian.....	3
<b>II. TINJAUAN PUSTAKA</b> .....	<b>5</b>
A. <i>Text Mining</i> .....	5
B. <i>Text Classification</i> .....	5
C. <i>Sentiment Analysis</i> .....	6
D. Twitter.....	6
E. <i>Application Programming Interface (API) Twitter</i> .....	11
F. <i>Word2Vec Model</i> .....	12
G. <i>Deep Learning</i> .....	13
H. Python.....	13
I. <i>Recurrent Neural Network (RNN)</i> .....	14
J. Protokol Kesehatan Pandemi Virus Corona.....	16
K. Virus Corona.....	16
L. Penelitian Sebelumnya.....	17

<b>III. METODE PENELITIAN .....</b>	<b>21</b>
A. Perangkat Penelitian .....	21
B. Waktu dan Tempat Penelitian.....	22
C. Diagram Alir Penelitian.....	22
D. Tahapan Penelitian .....	24
E. Menentukan Bobot Word2Vec dan Klasifikasi RNN .....	27
F. Pengukuran Kinerja .....	29
<b>IV. HASIL DAN PEMBAHASAN .....</b>	<b>31</b>
A. Analisis dan Klasifikasi Data Teks.....	31
B. Arsitektur Jaringan .....	36
C. Pengujian Parameter .....	37
D. <i>Training Data</i> .....	39
E. Evaluasi Data Uji.....	40
F. Visualisasi Kata pada Setiap Kelas .....	44
<b>V. SIMPULAN DAN SARAN .....</b>	<b>48</b>
A. Simpulan.....	48
B. Saran .....	49
<b>DAFTAR PUSTAKA .....</b>	<b>50</b>

## DAFTAR TABEL

Tabel	Halaman
1. Penelitian sebelumnya.....	17
2. Perangkat penelitian. ....	21
3. Jadwal penelitian. ....	22
4. <i>Confusion matrix</i> . ....	29
5. Data yang diberi label. ....	31
6. Hasil pengujian parameter RNN <i>Learning</i> . ....	37
7. Hasil pengujian parameter LSTM. ....	38
8. Kinerja klasifikasi model RNN <i>Learning</i> .....	41
9. Kinerja klasifikasi model <i>Naive Bayes</i> .....	43
10. Kinerja klasifikasi model LSTM.....	44

## DAFTAR GAMBAR

Gambar	Halaman
1. Halaman beranda Twitter.....	7
2. Halaman <i>followers</i> pada Twitter.....	7
3. Halaman <i>following</i> pada Twitter.....	8
4. Halaman profil pada Twitter.....	8
5. Halaman pesan pada Twitter.....	9
6. Halaman <i>trending topic</i> pada Twitter.....	9
7. Pencarian melalui <i>hashtag</i> pada Twitter.....	10
8. Contoh <i>mentions</i> pada Twitter.....	10
9. Halaman <i>like tweets</i> pada Twitter.....	11
10. Realtime Twitter Streaming API.....	12
11. Konsep dasar metode RNN <i>learning</i> dengan <i>looping</i> .....	15
12. Konsep dasar lanjutan metode RNN.....	15
13. Diagram alir penelitian.....	23
14. <i>Simple recurrent neural network</i> (Mikolov et al., 2010).....	28
15. <i>Wordlink</i> pada kelas positif.....	33
16. <i>Wordcloud</i> pada kelas positif.....	33
17. <i>Wordlink</i> pada kelas negatif.....	34
18. <i>Wordcloud</i> pada kelas negatif.....	34
19. <i>Stopword</i> bahasa Indonesia.....	35
20. Tahapan <i>cleansing</i> .....	36
21. Hasil <i>training</i> data RNN <i>Learning</i> .....	39
22. Hasil <i>training</i> data LSTM.....	40
23. <i>Confusion matrix</i> RNN <i>Learning</i> .....	41
24. <i>Confusion matrix</i> <i>Naive Bayes</i> .....	42
25. <i>Confusion matrix</i> LSTM.....	43



26. <i>Wordcloud</i> hasil pada kelas positif. ....	45
27. <i>Wordlink</i> hasil pada kelas positif. ....	45
28. <i>Wordcloud</i> hasil pada kelas negatif.....	46
29. <i>Wordlink</i> hasil pada kelas negatif. ....	46

## I. PENDAHULUAN

### A. Latar Belakang

Teknologi telah menjadi kebutuhan utama dalam kehidupan manusia terutama untuk memberikan informasi secara cepat dan jangkauan yang luas. Salah satu alat komunikasi digital yang sering digunakan adalah media sosial (Risnantoyo et al., 2020). Kegunaan komunikasi media sosial bagi para pengguna adalah memberikan opini, saran, kritik, atau *sharing* tentang yang terjadi pada seseorang. Pada saat ini, khalayak ramai selalu mengabadikan setiap kegiatan seperti foto maupun teks pada media sosialnya. Hal ini merupakan sebuah fenomena karena kemajuan yang besar dan cepat pada teknologi informasi. Salah satu aplikasi media sosial yang populer adalah Twitter.

Pada aplikasi Twitter, salah satu fiturnya adalah dapat melihat *tweets* atau ulasan tentang kegiatan seseorang dengan menambah *keyword*. Informasi yang diberikan melalui *keyword* tersebut dapat dilihat jika mengakses beranda Twitter atau fitur *search* sehingga informasi tersebut tersebar dengan cepat kepada para pengguna. Sebagai media informasi, saat ini ulasan pada Twitter tentang virus baru, yaitu virus corona menjadi berita utama hingga menarik perhatian *World Health Organization* (WHO). Virus corona adalah virus yang bernama *Novel Corona (2019-nCoV)* dengan gejala seperti penyakit *influenza* disertai sesak nafas dan *pneumonia*. Dampak yang diberikan oleh virus corona bukan hanya menyerang sektor kesehatan, tetapi juga menyerang sektor lainnya, seperti pariwisata, industri, bisnis, dan lainnya (Risnantoyo et al., 2020). Virus corona menyebar dengan cepat sehingga WHO menyebut penyakit virus ini sebagai pandemi. Pandemi ini terjadi hampir di seluruh dunia, sehingga untuk mencegah penyebaran dan menghindarkan diri dari resiko tertular virus corona, masyarakat harus menerapkan protokol kesehatan untuk beraktivitas sehari-hari. Berdasarkan Keputusan Menteri

Kesehatan RI Nomor HK.01.07/MENKES/382/2020 Tentang Protokol Kesehatan bagi Masyarakat di Tempat dan Fasilitas Umum dalam Rangka Pencegahan dan Pengendalian *Corona Virus Disease* 2019, protokol kesehatan yang dapat diterapkan diantaranya adalah untuk melakukan kegiatan di rumah saja, memakai pelindung diri atau masker apabila keluar rumah, mencuci tangan dengan sabun dan atau *handsanitizer*, serta menjaga jarak minimal satu meter dengan orang lain (Sitohang, 2020). Hal ini memunculkan berbagai *sentiment* dari masyarakat, baik *sentiment* positif, *sentiment* negatif, dan *sentiment* yang netral. Berdasarkan fenomena tersebut, dapat dilakukan riset suatu *sentiment analysis* melalui informasi pada Twitter dengan *keyword tweets* “dirumah saja”, ”pakai masker”, “jaga jarak”, dan “*stay safe*”.

*Sentiment analysis* atau sering disebut juga *opinion mining* adalah studi komputasi opini dari berbagai opini sekelompok orang, sentimen, emosi, dan penilaian terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik lainnya (Liu, 2015), Opini dan komentar dalam penelitian ini digunakan untuk melihat persepsi atau kecenderungan opini masyarakat terhadap protokol kesehatan untuk beraktivitas sehari-hari pada masa pandemi virus corona. *Sentiment analysis* pada aplikasi Twitter dilakukan dengan menggunakan *Application Programming Interface* (API) Key Twitter untuk mengambil opini atau komentar dari suatu forum atau *keyword tweets*. API Key Twitter adalah sekumpulan perintah, fungsi, dan protokol yang dapat digunakan developer dalam mengembangkan *software* sesuai dengan kebutuhannya masing-masing (Monarizqa et al., 2014).

*Sentiment analysis* pada penelitian ini menggunakan model *Word2Vec* sebagai fitur ekstraksi dan dipadukan dengan metode *Recurrent Neural Network* (RNN) *Learning*. Metode RNN digunakan untuk melakukan penelitian *sentiment analysis* pada persepsi masyarakat terhadap protokol kesehatan pada masa pandemi virus corona di aplikasi Twitter dengan data berupa teks. *Word vector* diusulkan sebagai cara baru untuk merepresentasikan kata di bidang pengolahan bahasa alami dan dibentuk dengan bantuan *Neural Network*. Teknik ini diusulkan dengan nama *Word2Vec* (Mikolov et al., 2013). Model *Word2Vec* ini mengolah kata yang

memiliki makna kata mirip tetapi hasil vektor yang dihasilkan juga mirip. Kemudian menggunakan metode *RNN Learning* yang merupakan salah satu model *Neural Network* dengan tidak bergantung pada aturan pemrosesan bahasa alami (Mikolov et al., 2010). Hal ini memudahkan dalam memprediksi konteks kata dengan cara melihat kedekatan sebuah kata dengan kata lain dengan posisi sebelum atau sesudah kata tersebut, sehingga pengenalan *sentiment analysis* memperoleh akurasi yang lebih tinggi berdasarkan teks yang diolah.

## **B. Rumusan Masalah**

Berdasarkan latar belakang masalah *sentiment analysis* terhadap pandemi virus corona yang terjadi saat ini, maka rumusan masalah pada penelitian ini adalah bagaimana pengklasifikasian *sentiment negatif*, *sentiment positif*, dan *sentiment* yang netral pada persepsi masyarakat terhadap protokol kesehatan virus corona di aplikasi Twitter menggunakan *Word2Vec Model* dan *Recurrent Neural Network (RNN) Learning*.

## **C. Batasan Masalah**

Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Pembuatan sistem dibangun menggunakan bahasa Python.
2. Model yang digunakan adalah *Word2Vec Model*.
3. Metode yang digunakan adalah metode *RNN Learning*.

## **D. Tujuan Penelitian**

Tujuan penelitian ini adalah penerapan metode *RNN Learning* dan *Word2Vec Model* untuk pengklasifikasian *sentiment analysis* untuk melihat persepsi atau kecenderungan opini masyarakat terhadap protokol kesehatan untuk beraktivitas sehari-hari pada masa pandemi virus corona di Twitter.

## **E. Manfaat Penelitian**

Manfaat dari penelitian adalah sebagai berikut.

1. Memberikan informasi untuk melihat kecenderungan persepsi masyarakat terhadap *sentiment analysis* pada persepsi masyarakat terhadap protokol kesehatan pada masa pandemi virus corona di aplikasi Twitter.
2. Memberikan kontribusi penelitian dalam kajian *sentiment analysis* pada Twitter.

## II. TINJAUAN PUSTAKA

### A. *Text Mining*

*Text mining* (penambangan teks) adalah penambangan data yang berupa teks dimana sumber data didapatkan dari beberapa dokumen, tujuannya untuk mencari kata-kata yang dapat mewakili isi dari dokumen tersebut sehingga dapat dilakukan analisis antar dokumen (Langgeni et al., 2010). *Text mining* berusaha untuk menemukan pola baru dari sekumpulan teks yang berjumlah besar, dengan kata lain, dapat menggali informasi dari sekumpulan dokumen menggunakan *tools* analisis yang ada dalam komponen-komponen *data mining*. Proses pokok yg terdapat pada *text mining* adalah *text preprocessing*, *text transformation*, *feature selection*, dan *pattern discovery* (Jacob et al., 2019).

### B. *Text Classification*

*Text classification* didefinisikan sebagai pengelompokkan teks secara otomatis ke satu atau beberapa kelas yang telah ditentukan sebelumnya (Darujati, 2016). Setelah itu model klasifikasi dibangun untuk memprediksi kategori teks baru. *Text classification* akan meniru bagaimana manusia memproses dan menyimpulkan makna dari teks. Bahasa alami mengandung semantik dan konsep abstrak yang sulit diterapkan dalam bahasa komputer. Misalnya, makna kata dapat berubah tergantung pada konteks saat digunakan. Masalah lain jika terdapat kesalahan ketik atau kesalahan ejaan dan singkatan. Untuk menangani ambiguitas perlu dijelaskan dalam pengklasifikasian. *Text classification* digunakan untuk melakukan peringkasan teks dengan tujuan mendapatkan makna dari beberapa dokumen teks sehingga teks hasil dapat menyampaikan informasi penting secara otomatis dan tidak lebih panjang dari sumber teks (Hong et al., 2019).

### C. *Sentiment Analysis*

*Sentiment analysis* atau sering disebut juga *opinion mining* adalah studi komputasi opini dari berbagai opini sekelompok orang, sentimen, emosi, dan penilaian terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik lainnya (Liu, 2015). Jika terdapat satu set dokumen teks berisi opini (sentimen) mengenai suatu objek, maka tujuan *opinion mining* adalah mengekstraksi atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen, selanjutnya dianalisis apakah komentar tersebut adalah komentar positif, komentar negatif, atau netral.

*Opinion mining* adalah bagian dari pekerjaan yang melakukan peninjauan berkaitan dengan perlakuan komputasional opini, sentimen, dan persepsi dari teks. Hal ini memungkinkan pengguna untuk mencari informasi tentang deteksi *flame* (bahasa yang sangat kasar) dalam *email* atau jenis komunikasi lainnya, persepsi produk baru, persepsi merk, dan manajemen reputasi (Thomas & C.A, 2018).

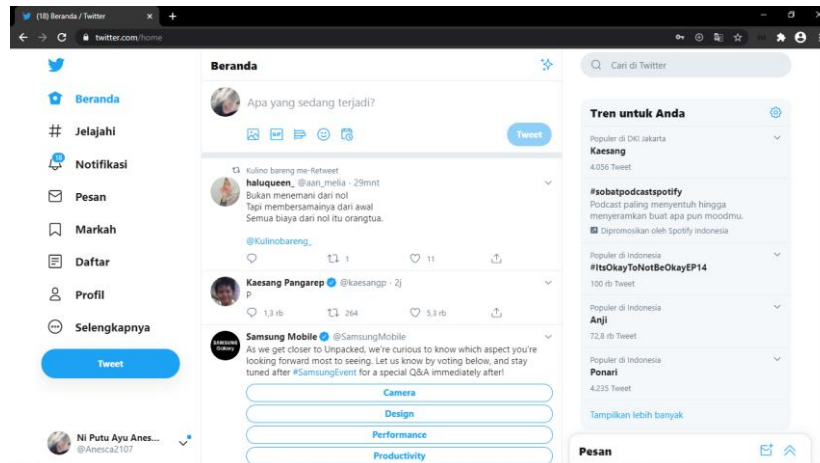
### D. **Twitter**

Twitter adalah sebuah aplikasi atau situs media sosial yang sedang populer pada saat ini. Seperti media sosial pada umumnya, Twitter menyediakan berbagai fitur agar pengguna dapat berinteraksi dengan pengguna lainnya kapan saja dan di mana saja. Twitter dapat diakses melalui *website* maupun aplikasi *mobile*. Menurut Chiang (2011), setelah Twitter diluncurkan pada Juli 2006, jumlah pengguna Twitter meningkat pesat. Pada September 2010, jumlah pengguna Twitter yang terdaftar mencapai sekitar 160 juta pengguna. Tidak seperti Facebook, LinkedIn, dan Myspace, Twitter merupakan sebuah jejaring sosial yang dapat digunakan sebagai sebuah graf berarah. Artinya, pengguna dapat mengikuti pengguna lain, tetapi pengguna kedua tidak perlu untuk mengikutinya kembali. Biasanya akun berstatus publik sehingga dapat diikuti tanpa persetujuan pemilik.

Fitur-fitur yang ada dalam Twitter (Twitter, 2020), antara lain:

## 1. Halaman Beranda (*Home*)

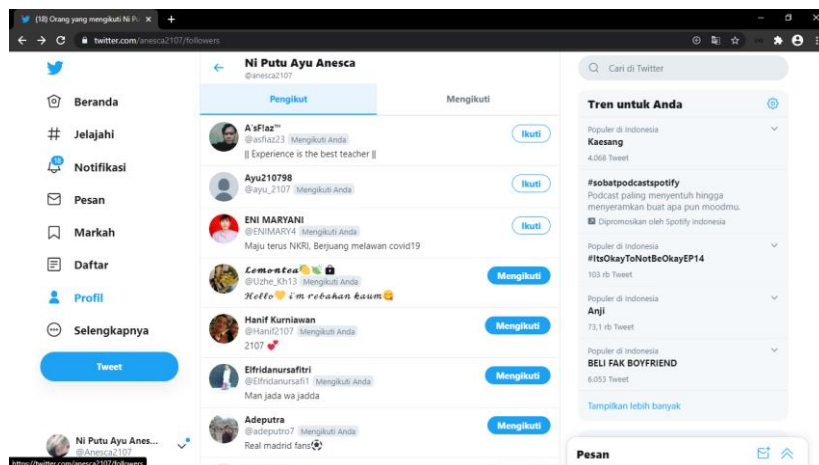
Pada halaman ini, pengguna dapat melihat *tweets* yang dikirim oleh pengguna yang menjadi teman atau pengguna yang diikuti (*following*). Sebagai contoh dapat dilihat pada Gambar 1.



Gambar 1. Halaman beranda Twitter.

## 2. Pengikut (*Followers*)

*Followers* adalah pengguna lain yang mengikuti aktivitas seseorang di Twitter, sehingga *tweets* yang dibuat oleh seseorang akan masuk ke dalam halaman utama pengikutnya. Sebagai contoh dapat dilihat pada Gambar 2.

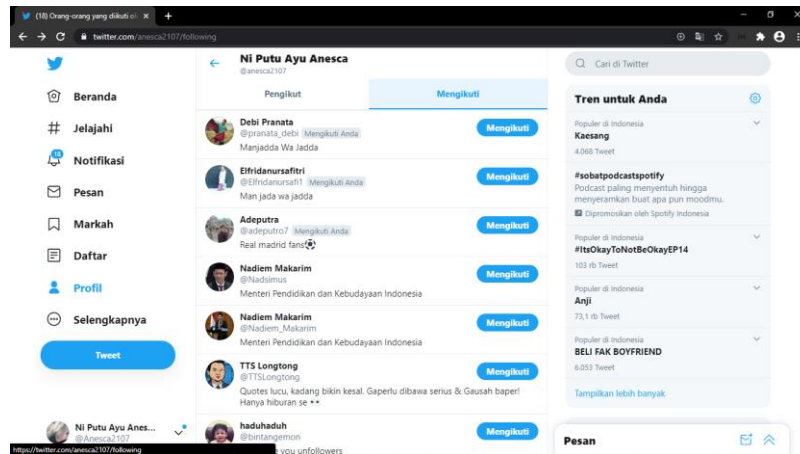


Gambar 2. Halaman *followers* pada Twitter.



### 3. Mengikuti (*Following*)

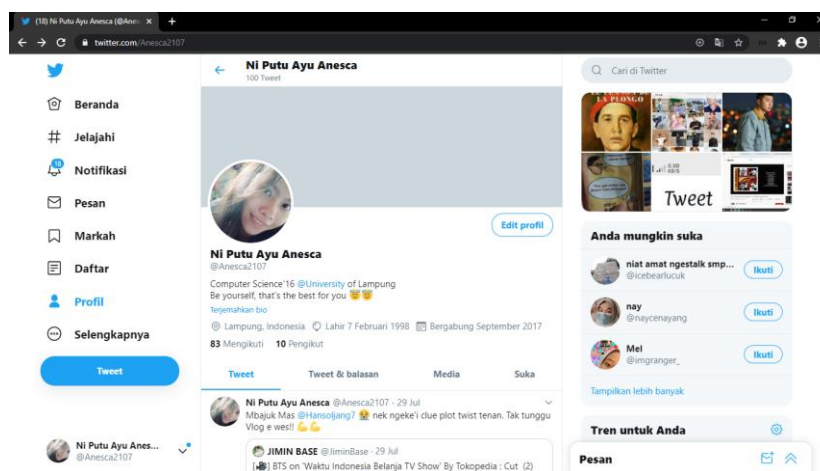
Kebalikan dari pengikut, *following* adalah mengikuti akun pengguna lain agar *tweets* yang dikirim oleh akun tersebut masuk ke dalam halaman utama. Sebagai contoh dapat dilihat pada Gambar 3.



Gambar 3. Halaman *following* pada Twitter.

### 4. Profil (*Profile*)

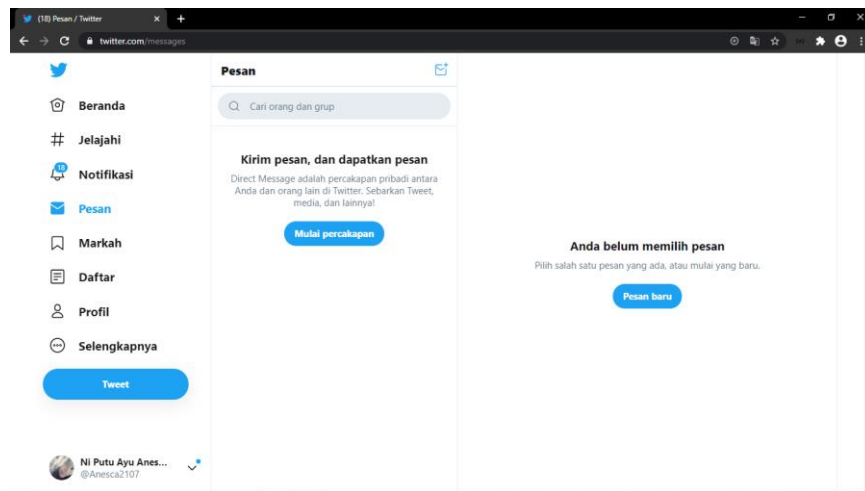
*Profile* adalah halaman yang bisa dilihat oleh semua orang mengenai profil atau data diri, serta *tweets* yang pernah dibuat. Sebagai contoh dapat dilihat pada Gambar 4.



Gambar 4. Halaman profil pada Twitter.

## 5. Pesan Langsung (*Direct Message*)

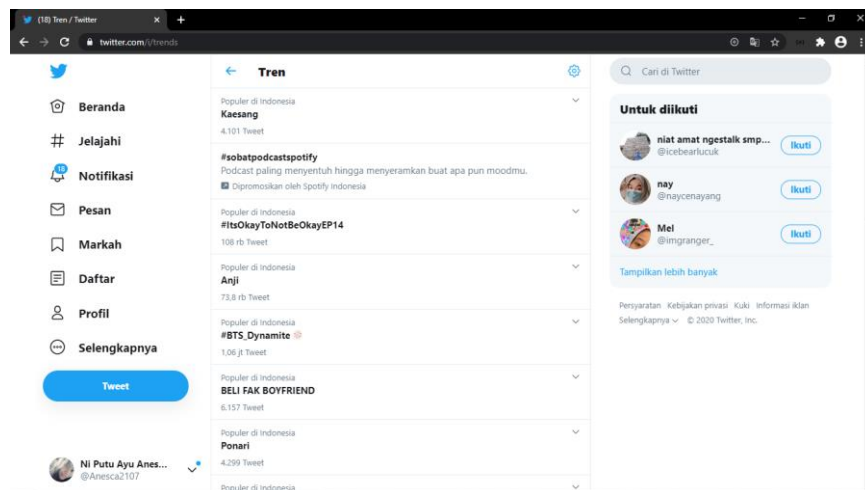
Pesan langsung adalah fitur untuk mengirimkan pesan kepada pengguna lain. Sebagai contoh dapat dilihat pada Gambar 5.



Gambar 5. Halaman pesan pada Twitter.

## 6. Topik Terkini (*Trending Topic*)

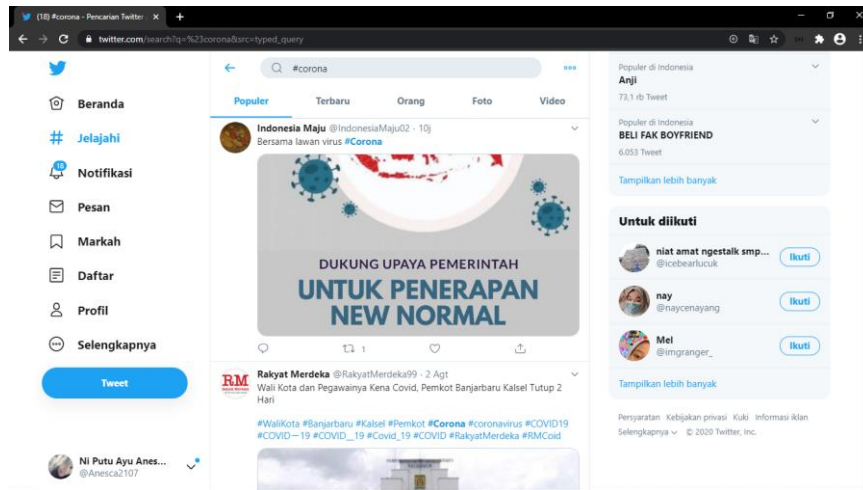
*Trending topic* adalah *tweets* yang sedang ramai dibicarakan oleh banyak pengguna pada waktu yang bersamaan dengan *keyword* yang sama. Sebagai contoh dapat dilihat pada Gambar 6.



Gambar 6. Halaman *trending topic* pada Twitter.

## 7. Hashtag

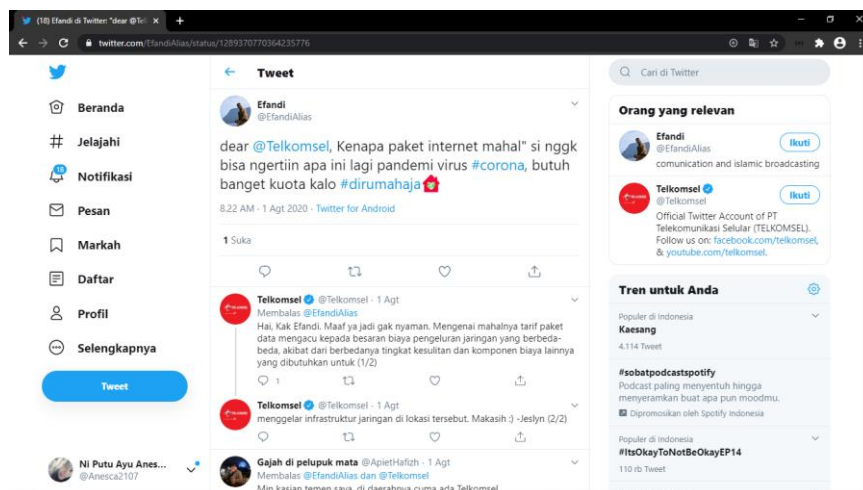
*Keyword* adalah simbol pagar (#) yang ditulis di depan topik tertentu agar pengguna lain dapat mencari *tweets* sejenis yang ditulis oleh pengguna lain. Sebagai contoh dapat dilihat pada Gambar 7.



Gambar 7. Pencarian melalui *hashtag* pada Twitter.

## 8. Mentions

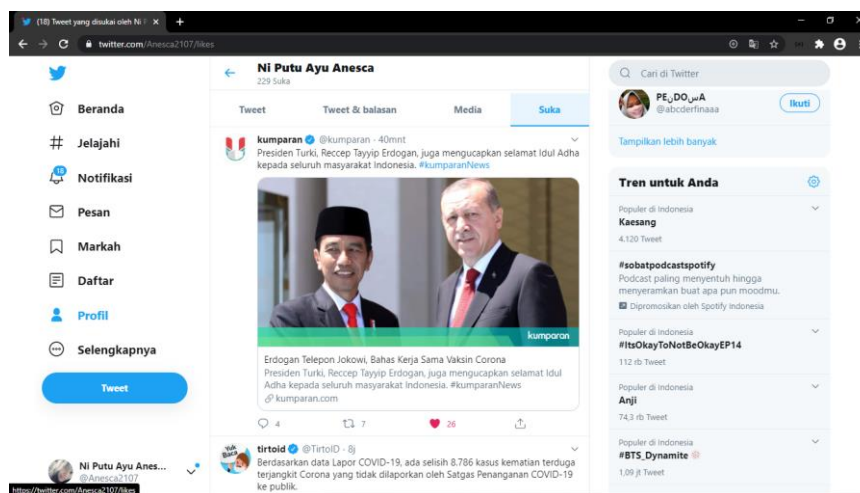
Biasanya, fitur ini adalah balasan *tweets* dari sebuah percakapan agar sesama pengguna dapat menandai akun lain. Sebagai contoh dapat dilihat pada Gambar 8.



Gambar 8. Contoh *mentions* pada Twitter.

## 9. Like Tweets

Fitur ini adalah menandai *tweets* pengguna lain sebagai favorit sehingga *tweets* tersebut tersimpan ke dalam fitur *like*. Sebagai contoh dapat dilihat pada Gambar 9.



Gambar 9. Halaman *like tweets* pada Twitter.

## E. Application Programming Interface (API) Twitter

API Twitter adalah sekumpulan URL (*Uniform Resource Locator*) untuk mengambil perintah, fungsi, dan parameter yang dapat digunakan dalam membangun *software* untuk sistem operasi tertentu. URL tersebut mengizinkan pengguna Twitter untuk membuat *update tweets*, mencari *tweets* berdasarkan *keyword* tertentu, dan fitur-fitur lainnya (Buslim et al., 2018). API mempunyai keunggulan dalam hal interaksi antar aplikasi, antara lain:

1. Portabilitas, *developer* yang menggunakan API dapat menjalankan programnya dalam sistem operasi apapun apabila sudah terpasang API tersebut.
2. Dalam hal *editing* dan pengembangan sebuah *software*, hal ini sangat penting karena API menggunakan bahasa yang lebih terstruktur dan mudah dimengerti daripada bahasa *system call*.

API Twitter yaitu sebuah aplikasi yang diciptakan oleh pihak Twitter untuk memudahkan *developer* lain mengakses informasi *web* Twitter tersebut dengan ketentuan dan syarat yang tertulis pada situs <http://dev.twitter.com/oauth>.

Terdapat dua jenis API Twitter, yaitu *REST API* dan *Streaming API*.

### 1. Twitter *REST API*

Twitter *REST API* terdiri dari Twitter *REST* dan Twitter *Search*. Twitter *REST* memberikan *core data* dan *core Twitter objects*. Di lengkapi dengan Twitter *Search* yang digunakan untuk melakukan pencarian mengenai suatu *instance* objek Twitter maupun mencari *trend*.

### 2. Twitter *Streaming API*

API ini biasanya digunakan untuk penggalian data, karena melalui API ini informasi didapatkan secara *realtime* dengan volume yang sangat tinggi. Twitter *Streaming API* menawarkan dua opsi *realtime* seperti pada gambar 10.

API	Category	Number of filters	Filtering operators	Rule management
<a href="#">statuses/filter</a>	Standard	400 keywords, 5,000 userids and 25 location boxes	Standard operators	One filter rule on one allowed connection, disconnection required to adjust rule
<a href="#">PowerTrack</a>	Enterprise	Up to 250,000 filters per stream, up to 2,048 characters each	Premium operators	Thousands of rules on a single connection, no disconnection needed to add/remove rules using Rules API

Gambar 10. *Realtime Twitter Streaming API*.

Pada Gambar 10, dapat dijelaskan sebagai berikut:

- Twitter *Streaming API Filter* adalah operasi *standard*, hanya dapat menyaring 400 *keywords*, 5000 pengguna, dan 25 lokasi dalam satu *rule* API pada satu koneksi yang diizinkan, pemutusan koneksi diperlukan untuk menyesuaikan *rule*.
- Twitter *Streaming API PowerTrack* adalah operasi berkualitas yang mampu mencapai 250.000 *filter* per aliran, masing-masing 2.048 karakter. Ribuan *rule* dalam satu koneksi dan tidak ada pemutusan koneksi yang diperlukan untuk menyesuaikan *rule* menggunakan *Rules API*.

## F. Word2Vec Model

Word2Vec *Model* adalah algoritme untuk mempelajari posisi kedekatan *semantic* antar kata dari sebuah teks masukan (Jang et al., 2019). Word2Vec *Model*

terdiri dari dua buah algoritme utama *word embedding*, yaitu: *Continuous Bag of Word* (CBOW) dan *Skip Gram*. Algoritme CBOW digunakan untuk melihat panjang tertentu dari sebuah kata pada dokumen masukan. Sebaliknya, algoritme *Skip Gram* digunakan untuk memprediksi konteks kata dengan cara melihat kedekatan sebuah kata dengan kata lain dengan posisi sebelum atau sesudah kata tersebut. Secara arsitektur, *Word2Vec Model* sebenarnya hanya sebuah jaringan syaraf tiruan yang tidak mempunyai banyak *hidden layer*, baik secara *node* dalam tiap *layer* maupun banyaknya *layer*.

### G. *Deep Learning*

*Deep Learning* adalah kemampuan komputer dalam mempelajari klasifikasi secara langsung dari sebuah data gambar, teks ataupun suara. Komputer dilatih menggunakan *dataset* berlabel dalam jumlah yang besar kemudian mengubah nilai data menjadi *feature vector* agar dapat mendeteksi pola input (Ranzato et al., 2015).

Metode ini adalah bagian dari *Neural Network*. *Deep Learning* mencakup banyak jaringan, diantaranya *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), *Deep Belief Network* (DBN), dan lain-lain. *Deep Learning* merupakan rekayasa perangkat lunak yang ditingkatkan, aksesibilitas daya komputasi dan data pelatihan, serta prosedur yang disempurnakan. Dasar dari penelitian *Deep Learning* adalah ilmu *neuron*, *neuron* membentuk jaringan yang digunakan sebagai cara mempelajari struktur model, kuantitas *layer*, dan jumlah *hidden layer* pada setiap lapisan (Ain et al., 2017).

### H. Python

Python adalah bahasa pemrograman tingkat tinggi yang didesain untuk pembacaan kode dan sintaks dengan tujuan untuk memudahkan para *programmer* mengekspresikan konsep ke dalam beberapa baris kode. Python dapat diaplikasikan ke aplikasi *web*, aplikasi *desktop*, *server* jaringan, *machine learning*, alat media, dan lain-lain (Ain et al., 2017). Bahasa pemrograman Python mendukung konsep pemrograman berorientasi objek (*object oriented programming*), sehingga terdapat

berbagai macam *library* dan *framework* yang dikembangkan yang berguna untuk melakukan analisis data, yaitu:

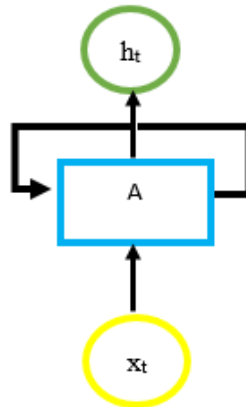
- *Scikit-Learn*, adalah *library* Python yang biasa digunakan untuk membangun sebuah program tentang *machine learning*. *Scikit-learn* menyediakan *supervised* dan *unsupervised algorithm* untuk Python (Pedregosa et al., 2011).
- *Scrapy* adalah *framework* yang digunakan untuk melakukan *crawling* pada sebuah situs *web* dan ekstraksi suatu data yang dapat digunakan untuk proses *data mining* (Pedregosa et al., 2011).
- Keras, bahasa pemrograman Python memiliki *library* bernama keras. Keras adalah *library neural network* tingkat tinggi yang mampu berjalan di atas CNTK, Theano, atau TensorFlow. Fitur yang disediakan dalam *library* ini digunakan untuk mempermudah pengembangan *Deep Learning* lebih dalam (Chollet, 2020).

### I. **Recurrent Neural Network (RNN)**

*Recurrent Neural Network* (RNN) adalah salah satu model *Neural Network* yang tidak bergantung pada aturan pemrosesan bahasa alami. RNN dapat menyimpan semua nilai input yang dilihat oleh *network* dan juga nilai input yang berlaku, nilai yang disembunyikan di setiap *layer network* bergantung pada semua input yang terlihat sebelumnya (Mikolov et al., 2010). RNN masuk ke dalam kategori *Deep Learning* karena data diproses secara otomatis dan tanpa pendefinisian fitur. RNN dapat menggunakan memori untuk memproses urutan input (Sari et al., 2021).

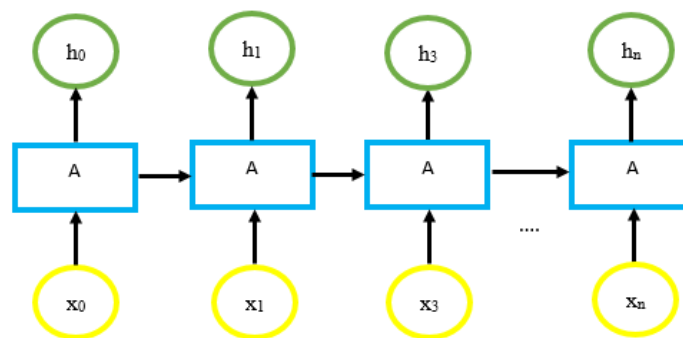
Penerapan dalam metode RNN *Learning* dan *Word2Vec Model* menggunakan konsep *back propagation*. *Back propagation* merupakan salah satu dari pembelajaran *machine learning*. *Back propagation* adalah proses membagikan *error* yang ada ke *layer-layer* sebelumnya. Algoritme ini memiliki dua fase untuk memproses data. Fase pertama yaitu vektor input dikirim ke *layer* input, lalu dilanjutkan ke *layer* proses. Kemudian nilai *output* ditemukan pada *layer output*. Fase kedua, jika nilai *output* berbeda dari nilai target, *error* nilai *output* akan dihitung dan dikembalikan ke lapisan input. Dengan *error* sekecil mungkin, sistem dapat bekerja dengan baik. Jadi, kondisi ketika akan melakukan sesuatu, maka akan

memprediksi kejadian yang selanjutnya (Prasetyo, 2012). Gambar 11 adalah ilustrasi dari konsep dasar RNN *Learning* dengan *Looping*.



Gambar 11. Konsep dasar metode RNN *learning* dengan *looping*.

Menurut Malik (2018), konsep dasar pada Gambar 11 menjelaskan bahwa  $x_t$  adalah *layer input* yang dimasukkan ke *layer A* (*layer proses*) maka akan menghasilkan *layer output*  $h_t$ . Selanjutnya, *layer output* melakukan *looping* sebagai *layer input*. *Looping* ini bermanfaat untuk menganalisis kemungkinan fakta-fakta dari *layer output* ( $h_t$ ). Uraian yang lebih tepat dari *looping* tersebut dapat dilihat pada Gambar 12.



Gambar 12. Konsep dasar lanjutan metode RNN.

Rantai pada Gambar 12 menjelaskan bahwa RNN berkaitan erat dengan daftar dan *sequence*. *Layer x* sebagai *layer input*, *layer A* sebagai *layer proses*, dan  $h$  sebagai *layer output*.  $x_0$  adalah sebuah *input* kata yang telah diubah menjadi vektor. Kemudian, input diproses melalui *layer A* sehingga menghasilkan *output*  $h_0$ . Dari *output*  $h_0$ , vektor menjadi input kembali sebagai  $x_1$  sebagaimana proses  $x_0$ , begitu



seterusnya hingga proses  $x_n$ . Arsitektur jaringan syaraf alami sangat digunakan untuk data tersebut.

## **J. Protokol Kesehatan Pandemi Virus Corona**

Pandemi virus corona terjadi hampir di seluruh dunia. Penyebarannya terjadi dengan sangat cepat, sehingga untuk mencegah penyebaran dan menghindarkan diri dari resiko tertular virus corona, masyarakat harus menerapkan protokol kesehatan untuk beraktivitas sehari-hari. Berdasarkan Keputusan Menteri Kesehatan RI Nomor HK.01.07/MENKES/382/2020 Tentang Protokol Kesehatan bagi Masyarakat di Tempat dan Fasilitas Umum dalam Rangka Pencegahan dan Pengendalian *Corona Virus Disease 2019*, protokol kesehatan yang dapat diterapkan diantaranya adalah untuk melakukan kegiatan di rumah saja, mencuci tangan dengan sabun dan atau *handsanitizer*, serta menjaga jarak minimal satu meter dengan orang lain (Kementerian Kesehatan, 2020).

## **K. Virus Corona**

Menurut Sitohang (2020), virus corona adalah virus yang bernama *Novel Corona (2019-nCoV)* yang muncul pada akhir tahun 2019. Ciri-ciri orang yang terinfeksi virus ini gejalanya adalah demam, flu, batuk yang disertai sesak nafas, dan *pneumonia*. Saat ini, belum tersedia vaksin untuk mengobati *2019-nCoV*. Virus ini tersebar oleh orang yang memiliki riwayat perjalanan ke kota Wuhan, China. Maka, hal ini dinyatakan WHO sebagai keadaan darurat dunia karena hingga tanggal 25 Maret 2020, WHO meng-*update* laporan ke-65 yang melaporkan bahwa terdapat 414.179 kasus virus corona dan 18.440 kematian di dunia. Hal ini menjadi topik utama yang ramai dalam media informasi karena penularannya yang sangat cepat (Sitohang, 2020). Pada bulan September 2020, kasus virus corona mencapai 29.700.000 terkonfirmasi positif dan 937.000 kematian di 216 negara. Informasi ini menjadi perbincangan publik di media sosial dengan *keyword* “virus corona” sehingga kerap menempati kata populer.

## L. Penelitian Sebelumnya

Penelitian ini mengacu pada penelitian-penelitian yang terkait sebelumnya. Hal tersebut digunakan sebagai rujukan dan perbandingan pada metode yang digunakan serta hasil yang dicapai pada penelitian ini. Ringkasan penelitian sebelumnya diuraikan pada Tabel 1.

Tabel 1. Penelitian sebelumnya.

No	Penulis	Judul Penelitian	Tahapan Penelitian	Hasil Penelitian
1.	(Sitohang, 2020)	Analisis Sentimen Pembangunan Infrastruktur di Indonesia dengan <i>Automated Lexicon Word2Vec</i> dan <i>Naive-Bayes</i>	- Pengujian menggunakan 150 data latih sentimen positif dan sentimen negatif, - Pengujian menggunakan 50 data uji sentimen positif dan sentimen negatif. - Pengujian terhadap metode Word2Vec menggunakan lima parameter, yaitu <i>windows size, learning rate, hidden neuron, Top-N kata,</i> dan <i>epoch.</i>	Penggunaan metode <i>Naive-Bayes</i> memiliki nilai <i>precision</i> 0,36, <i>recall</i> 0,818, <i>f-measure</i> sebesar 0,5 dan akurasi sebesar 64%. Semakin banyak variasi dan jumlah data latih yang digunakan untuk membuat <i>automated lexicon</i> , maka akan semakin akurat hasil prediksi kelas data uji.

Lanjutan Tabel 1. Penelitian sebelumnya.

No	Penulis	Judul Penelitian	Tahapan Penelitian	Hasil Penelitian
2.	(Hassan, 2017).	<i>Sentiment Analysis with Recurrent Neural Network and Unsupervised Neural Language Model</i>	- Menggunakan aspek <i>LSTM</i> dan aspek <i>RNN</i> yang mengontrol besar <i>current input</i> dalam menciptakan memori baru maupun memori sebelumnya, dan bagian yang penting dalam menghasilkan <i>output</i> .	Model ini memiliki potensi untuk mengatasi beberapa kelemahan dalam metode tradisional. Misalnya, model <i>bag-of-words</i> dan <i>n-gram</i> , maka pesan dan informasi dari kata tersebut menghilang
3.	(Thomas & C.A, 2018).	<i>Sentimental Analysis Using Recurrent Neural Network</i>	- Mengacu pada algoritme <i>Recurrent Neural Network-Long Short Term Memory (RNN-LSTM)</i> - <i>Dataset</i> dimasukkan ke <i>neural model</i> . - Setelah model dilakukan dan hasil <i>Batch Accuracy</i> diperoleh, kalimat sebagai input memprediksi <i>sentiment</i> dengan akurasi yang lebih besar.	Rata-rata akurasi yang diperoleh oleh model ini adalah 60%. Keakuratan hasil masih bisa ditingkatkan jika memiliki <i>dataset</i> yang lebih banyak. <i>Deep Learning</i> menggunakan <i>RNN</i> juga dapat digunakan untuk analisis <i>sentimental domain</i> untuk bahasa lain.

Lanjutan Tabel 1. Penelitian sebelumnya.

No	Penulis	Judul Penelitian	Tahapan Penelitian	Hasil Penelitian
4.	(Lestari et al., 2019).	Rekomendasi Lagu berdasarkan Lirik dan <i>Genre</i> Lagu menggunakan Metode <i>Word Embedding</i> (Word2Vec).	<ul style="list-style-type: none"> <li>- Pengujian menggunakan 25 lirik Query.</li> <li>- Nilai batas K yang digunakan untuk menghitung nilai evaluasi <i>precision</i> dan <i>recall</i> adalah 10 (<i>precision</i> 10).</li> <li>- Parameter-parameter yang digunakan yaitu <i>learning rate</i>, <i>hidden neuron</i>, <i>window size</i>, <i>epoch</i>, Top-N kata TF-IDF dan Top-N kata Word2Vec.</li> </ul>	<p>Word2Vec dapat menghitung tiap kata pada tiap lirik lagu yang ada pada <i>dataset</i> dan menghitung kedekatan kata tertinggi dengan kata-kata lain. Proses pembobotan tersebut dilakukan hingga <i>epoch</i> yang telah ditetapkan atau hingga konvergen. Nilai TF-IDF lagu dihitung untuk pencarian nilai kata tertinggi. Hasil dari TF-IDF lagu yang terbaik adalah sebanyak 10 kata dengan skor <i>precision</i> 10 sebesar 0.592 serta nilai <i>Mean Average Precision</i> sebesar 0.7161.</p>

Dari beberapa penelitian sebelumnya, tingkat akurasi menggunakan algoritme *Recurrent Neural Network* (RNN) memiliki tingkat akurasi sebesar 80%. Metode ini dapat mengatasi beberapa kelemahan metode lain seperti metode *bag-of-words*

yang dapat menghilangkan arti kata tersebut apabila telah dilakukan *tokenizing*. Penggunaan *Word2Vec model* pada penelitian yang lain apabila semakin banyak variasi dan jumlah data latih yang digunakan untuk membuat *automated lexicon*, maka akan semakin akurat juga hasil prediksi kelas data uji. Maka, untuk penelitian ini melakukan hal serupa dengan penelitian-penelitian sebelumnya yaitu melakukan *sentiment analysis* menggunakan model *Word2Vec* dan metode *RNN* namun dengan data yang berbeda. Data yang digunakan dalam penelitian ini adalah data teks pada Twitter dengan *keyword tweets* “dirumah saja”, “jaga jarak”, “pakai masker”, dan “*stay safe*”.

### III. METODE PENELITIAN

#### A. Perangkat Penelitian

Sebagai penunjang proses pembuatan sistem, dibutuhkan perangkat penelitian berupa *hardware* dan *software* yang ditunjukkan pada Tabel 2.

Tabel 2. Perangkat penelitian.

No	Perangkat	Spesifikasi	Kegunaan	Jumlah
1.	Laptop Acer Aspire E 14	<i>Processor</i> Intel Core i5-8250U 1.6GHz <i>up to</i> 3.4GHz, RAM 8 GB DDR4 <i>Memory</i> , 1TB HDD	Perangkat keras untuk membuat dan menguji aplikasi.	1
2.	Python	Aplikasi Anaconda based on Python	Bahasa pemrograman untuk membuat aplikasi.	1
3.	Windows 10	x64 <i>Bit</i>	Sistem operasi	1

## B. Waktu dan Tempat Penelitian

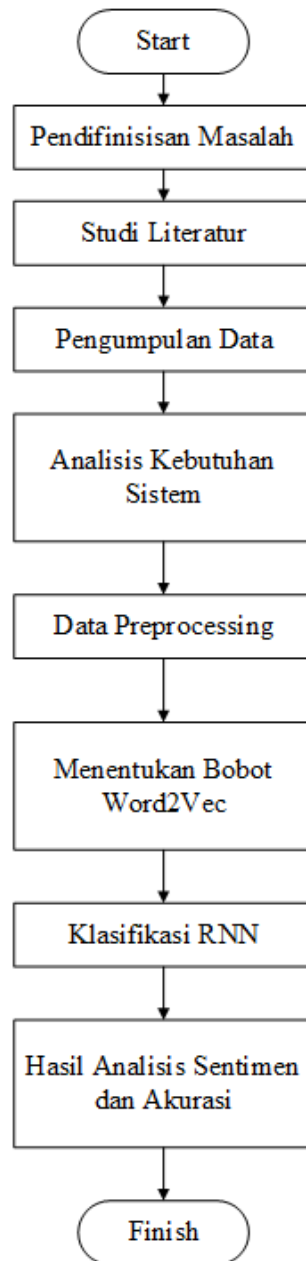
Penelitian ini dilaksanakan pada semester ganjil 2020/2021. yang bertempat di Jurusan Ilmu Komputer Universitas Lampung. Penelitian ini dibagi menjadi beberapa tahap sesuai metode penelitian dengan jadwal yang ditunjukkan pada Tabel 3.

Tabel 3. Jadwal penelitian.

Jadwal Penelitian	Bulan ke-									
	10	11	12	1	2	3	4	5		
Pendefinisian Masalah	■	■								
Studi Literatur		■	■							
Pengumpulan Data			■	■						
Analisis Kebutuhan Sistem					■	■				
Data <i>Preprocessing</i>						■	■	■	■	
Pengolahan Data <i>Preprocessing</i> dan Implementasi									■	■

## C. Diagram Alir Penelitian

Penelitian ini melalui beberapa tahapan yang diilustrasikan pada Gambar 13.



Gambar 13. Diagram alir penelitian.

Secara umum penelitian ini memiliki beberapa tahapan. Tahapan penelitian ini mencakup pendefinisian masalah, studi literatur, pengumpulan data, analisis kebutuhan sistem, data *preprocessing*, pengolahan data dan implementasi, serta penarikan simpulan.



## **D. Tahapan Penelitian**

### **1. Pendefinisian Masalah**

Pendefinisian masalah pada penelitian ini adalah bagaimana membuat *sentiment analysis* Twitter terkait *tweets* terhadap protokol kesehatan untuk beraktivitas sehari-hari pada masa pandemi virus corona menggunakan model *Word2Vec* dan metode *Deep Learning: Recurrent Neural Network (RNN)*. Indikator yang dapat diukur pada tahap ini adalah sejauh mana akurasi yang dapat diberikan oleh metode *RNN Learning*. Tingkat akurasi dapat diketahui dengan cara mengukur kinerja menggunakan *confusion matrix*, yaitu menghitung akurasi, presisi, dan *recall* dari data uji maupun *training* yang *positif* serta *negatif*.

### **2. Studi Literatur**

Tahapan studi literatur pada penelitian ini adalah dengan studi pustaka. Studi pustaka yaitu mencari, mempelajari, dan menggunakan berbagai literatur baik berupa buku, jurnal, *e-book*, atau literatur yang berkaitan dengan *Word2Vec model* dan metode *RNN Learning*, terutama dengan pembahasan *sentiment analysis*. Indikator yang dapat diukur adalah terkumpulnya literatur *RNN* yang dapat digunakan untuk menunjang penelitian ini.

### **3. Pengumpulan Data**

Tahap awal dalam pengumpulan data adalah mencari *dataset* yang cocok untuk penelitian. Indikator yang dapat diukur adalah terkumpulnya *dataset* dari aplikasi Twitter. *Dataset* yang digunakan dalam penelitian ini berjumlah 2000 data yang didapatkan dari API Twitter. API Twitter dapat dimanfaatkan untuk akan mengakses *server* Twitter lalu mengambil berbagai *tweets*. Data dari berbagai macam *tweets* pada media sosial Twitter dapat digunakan sebagai topik penelitian di bidang *data mining*. Dengan API Twitter, penulis mengambil berbagai macam *tweets* dengan *keyword tweets* “dirumah saja”, “pakai masker”, “jaga jarak”, dan “*stay safe*”. Saat ini, virus corona menjadi *trending topic* di Twitter karena penyebarannya yang sangat cepat, sehingga beberapa protokol kesehatan diterapkan pada masa pandemi virus corona. Oleh karena itu, banyak orang memberi ulasan yang *positif*, *negatif*, dan *netral* ke dalam sebuah *tweets*. Teknik

ini mengikuti penelitian sebelumnya oleh Kouloumpis et al., (2011), yaitu menggunakan kata-kata yang bermakna sentimen sebagai penanda sentimen pada *tweets* tersebut.

#### 4. Analisis Kebutuhan Sistem

Tahapan analisis kebutuhan sistem yaitu melakukan identifikasi terhadap sistem yang dibangun untuk melakukan *sentiment analysis* menggunakan metode RNN serta *Word2Vec model*. Indikator yang dapat diukur adalah terbentuknya *support system* untuk melakukan *sentiment analysis*.

##### a. Analisis Kebutuhan Data dan Informasi

Analisis kebutuhan data dan informasi untuk mengetahui data dan informasi yang dibutuhkan oleh sistem yang dibuat. Selanjutnya, digunakan untuk menganalisis hasil *output* dari pengguna tersebut berdasarkan hasil analisis kebutuhan pengguna, kebutuhan data, dan informasi.

##### b. Kebutuhan Fungsional

Kebutuhan fungsional merupakan kebutuhan sistem yang berhubungan dengan proses input dan *output* pada sistem. Dengan menganalisis kebutuhan fungsional, dapat diketahui fungsi-fungsi apa saja yang dapat dilakukan sistem informasi yang dibangun. Spesifikasi kebutuhan fungsional dari penelitian ini antara lain:

- Sistem dapat melakukan *text preprocessing* pada data *tweets*.
- Sistem dapat melakukan klasifikasi untuk menghitung probabilitas tertinggi, sehingga dapat membedakan data *tweets* yang berisi penilaian positif, negatif, dan netral.
- Sistem memiliki fitur untuk menampilkan hasil berupa informasi dari *tweets* yang diolah.

##### c. Kebutuhan Non-fungsional

Kebutuhan non-fungsional merupakan kebutuhan yang berada di luar fungsi-fungsi yang dapat dilakukan sistem. Dengan kata lain, yaitu kebutuhan yang berhubungan dengan batasan lingkungan sistem, kinerja sistem, keandalan sistem,

keamanan sistem, dan lain-lain. Kebutuhan non-fungsional dari sistem ini yaitu data *tweets* lebih mudah untuk diolah apabila menggunakan bahasa Indonesia yang baik dan benar.

## 5. **Data Preprocessing**

Tahapan ini merupakan tahap mengolah *dataset* yang telah didapat. Pada bagian ini, dibangun *word embedding* atau pengubahan kata menjadi sebuah vektor. Maka, data disiapkan agar dapat dianalisis dan diklarifikasi. Data *preprocessing* digunakan untuk menganalisis data yang dibutuhkan dalam proses *RNN Learning*. Tahapan *preprocessing* digunakan karena sistem tidak dapat membaca dokumen teks. Struktur pada dokumen teks tidak teratur sehingga diperlukan mengubah teks menjadi sebuah angka yang terstruktur agar dapat dikenali oleh sistem. Berikut adalah beberapa proses data *preprocessing* agar data sesuai dengan kebutuhan, yaitu:

### a. *Tokenizing*

*Tokenizing* adalah memisahkan setiap kata yang terdapat pada suatu dokumen. Dengan kata lain, yaitu memotong *string* dari setiap kata yang menyusunnya. Untuk melakukan pemisahan kata pada proses *tokenizing* biasanya berdasarkan oleh karakter spasi pada dokumen. Maka, dalam melakukan *tokenize*, ulasan-ulasan yang ada dipisahkan menjadi kalimat-kalimat terlebih dahulu. Kemudian, kalimat-kalimat tersebut dipisahkan lagi menjadi kata-kata. Sebagai contoh, misalnya terdapat kalimat “*Are you sure?*”, maka *tokenize*-nya adalah “*Are*”, “*you*”, “*sure*”, “*?*”.

### b. *Cleansing*

*Cleansing* adalah penghilangan tanda baca atau karakter yang tidak digunakan. Setelah itu, kata yang telah di-*tokenize* diubah menjadi huruf kecil (*lowercase*). Sistem mengumpulkan kata yang telah di-*tokenize* dan disimpan di dalam *database* untuk proses selanjutnya.

c. *Filtering*

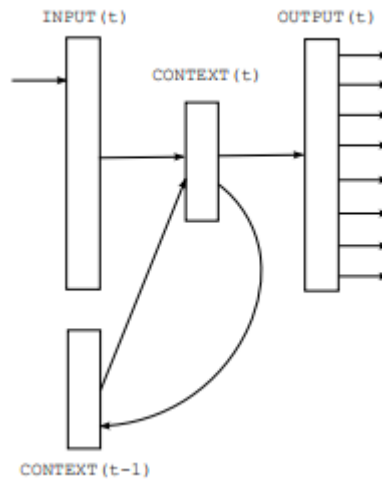
*Filtering* adalah salah satu proses *preprocessing* pada *text mining*, yaitu mengambil kata-kata yang penting saja atau biasa disebut dengan menghilangkan *stopword*. *Stopword* merupakan proses menghilangkan kata-kata yang paling sering muncul tetapi tidak memiliki pengaruh dan tidak dibutuhkan oleh pengguna dalam suatu dokumen. Daftar *stopword* yang dibuat digunakan sebagai proses acuan. Contoh *stopword list* ini berisi kata-yang sering digunakan, seperti: ke, dan, walaupun, tapi, yang, itu, ini, di sana, di sini, atau, maka, dari, dan sebagainya.

**E. Menentukan Bobot *Word2Vec* dan Klasifikasi RNN**

*Word2Vec* (*word embedding*) adalah salah satu model pengembangan dalam komputasi untuk merepresentasikan kata menjadi sebuah *vector* atau *array* bilangan riil (Utomo, 2015). Hasil dari *word embedding* dapat digunakan ke dalam ruang *embedding*. Kata-kata dengan makna yang serupa cenderung mempunyai jarak kedekatan yang kecil dalam satu bidang yang sama di ruang *embedding*. Maka, hal ini dapat menyebabkan kesamaan semantik antar kata. Sifat keragaman kata sangat penting dan harus diperhatikan dalam pengujian-pengujian statistik agar penarikan kesimpulan menjadi lebih akurat.

Vektor memudahkan pola linguistik dalam mengatur kata-kata. Pada dasarnya, *Word2Vec* dapat dilakukan menggunakan *neural network* atau faktorisasi matriks. Terdapat dua model prediksi RNN yang efisien untuk *Word2Vec*, yaitu *Continuous Bag-of-Words* model (CBOW), dan *Skip-Gram* model (SG). Model CBOW merupakan model yang memprediksi kata target. Sebagai contoh, terdapat kata-kata konteks yaitu: Lelaki itu sedang “...” topi. Maka, kata target yang ada dalam “...” adalah “memakai”. Sebaliknya, model *Skip Gram* yang memprediksi kata-kata yang konteks sehingga dapat diberi kata target. Secara statistik, model CBOW efektif untuk kumpulan *dataset* yang sedikit karena mudah memprediksi seluruh konteks. Namun, model SG memprediksi sepasang konteks-target sebagai proses baru dan lebih baik untuk *dataset* yang lebih besar (Zhang et al., 2018).

Proses *Word2Vec* pada penelitian ini menggunakan algoritme RNN untuk mengklasifikasikan sentimen. Modul RNN yang digunakan dalam penelitian ini adalah *simple recurrent neural network*. Menurut Mikolov et al., (2010), tujuan pemodelan bahasa statistik digunakan untuk memprediksi kata berikutnya dari teks yang konteks. Model untuk pengenalan kata atau sistem terjemahan dibuat dari data dalam jumlah yang besar. Gambar 14 menggambarkan konsep dari *simple recurrent neural network*.



Gambar 14. *Simple recurrent neural network* (Mikolov et al., 2010).

*Simple recurrent neural network* menggunakan koneksi yang berulang atau disebut dengan jaringan Elman, sehingga informasi dapat berputar secara berulang-ulang di dalam jaringan ini untuk waktu yang lama. Jaringan memiliki *layer* input  $x$ , *hidden layer*  $s$  (*layer context*) dan *layer output*  $y$ . Input ke jaringan pada waktu  $t$  adalah  $x(t)$ , *output* dilambangkan sebagai  $y(t)$ , dan status *hidden layer* dilambangkan sebagai  $s(t)$ . Vektor *input*  $x(t)$  dibentuk oleh vektor penggabungan  $w$  (kata saat ini), dan *output* dari *neuron* pada *layer context*  $s$  dibentuk pada waktu  $t - 1$ . *Layer input*, *hidden layer*, dan *layer output* dihitung menggunakan persamaan (1) sampai dengan persamaan (5) sebagai berikut:

$$x(t) = w(t) + s(t - 1) \quad (1)$$

$$s_j(t) = f \left( \sum_i x_i(t) u_{ji} \right) \quad (2)$$

$$y_k(t) = g \left( \sum_j s_j(t) u_{kj} \right) \quad (3)$$

dimana  $f(z)$  sebagai fungsi aktivasi sigmoid dengan rumus:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

dan  $g(z)$  sebagai fungsi *softmax*:

$$g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}} \quad (5)$$

Jumlah data latih harus sama dengan jumlah *layer context* dan *hidden layer*. *Output layer*  $y(t)$  adalah distribusi probabilitas dari kata berikutnya yang diberikan oleh kata sebelumnya  $w(t)$  dan *context*  $s(t - 1)$ . Fungsi *softmax* digunakan untuk memastikan distribusi probabilitas ini valid, yaitu untuk sembarang kata  $m$  dengan fungsi  $y_m(t) > 0$  dan  $\sum_k y_k(t) = 1$  (Mikolov et al., 2010).

## F. Pengukuran Kinerja

Pengukuran kinerja yang digunakan adalah *confusion matrix* atau tingkat *error matrix*. Parameter yang digunakan untuk mengukur kinerja RNN dalam *confusion matrix* yaitu TP (*True Positif*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negatif*) yang dapat dilihat pada Tabel 4 (Patel dan Tiwari, 2019).

Tabel 4. *Confusion matrix*.

	Label Positif Prediksi	Label Negatif Prediksi
Label Positif Aktual	TP	FN
Label Negatif Aktual	FP	TN

Untuk menghitung tingkat akurasi, *recall*, dan presisi adalah merujuk pada persamaan (6) sampai dengan persamaan (8) berikut.

## a. Akurasi

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Persamaan (6) adalah perbandingan antara kelas yang diprediksi dengan benar oleh sistem klasifikasi terhadap seluruh data untuk mengevaluasi kinerja model dengan menggunakan klasifikasi biner.

## b. Recall

$$\frac{TP}{TP + FN} \quad (7)$$

Persamaan (7) adalah perbandingan antara kelas yang diprediksi dengan benar oleh sistem terhadap total data yang ada pada suatu kelas, biasanya disebut sebagai sensitivitas.

## c. Presisi

$$\frac{TP}{TP + FP} \quad (8)$$

Persamaan (8) adalah perbandingan antara kelas yang diprediksi dengan benar oleh sistem terhadap total data yang telah diklasifikasi oleh suatu kelas, biasanya disebut sebagai nilai prediksi positif.

## V. SIMPULAN DAN SARAN

### A. Simpulan

Pada penelitian ini, dapat ditarik simpulan bahwa ulasan *tweet* berjumlah 2000 data. Data tersebut dibagi menjadi dua kelas, yaitu kelas negatif berjumlah 800 ulasan dan kelas positif berjumlah 1200 ulasan. Ulasan pada kelas negatif menunjukkan keluhan maupun kekecewaan dari *user* pada aturan yang diterapkan oleh pemerintah maupun tidak taatnya masyarakat terhadap protokol kesehatan. Kemudian dalam kelas positif menghasilkan ulasan yang mengandung kepuasan dari *user* dengan adanya protokol kesehatan virus corona yang layak diterapkan oleh masyarakat.

Pengujian dilakukan menggunakan 3 model, diantaranya yaitu model RNN *Learning*, *Naive Bayes* dan LSTM sebagai pembanding akurasi yang terbesar.

1. Model RNN *Learning*, dengan pengaturan parameter *W2V\_Window* sebesar 8, *W2V\_Epoch* sebesar 32, dan *W2V\_Min\_Count* sebesar 10, *neuron* sebesar 200, dan menggunakan aktivasi *sigmoid* menghasilkan akurasi bernilai 66%. Kemudian, proses *training* pada pengujian ini menggunakan sebanyak 10 epoch dan 32 batch size. Nilai *precision* yang didapatkan dari model RNN pada penelitian ini, untuk *sentiment* negatif sebesar 54% dan untuk *sentiment* positif sebesar 67%. Kemudian untuk nilai *recall* yang didapatkan pada *sentiment* negatif yaitu sebesar 11% dan untuk *sentiment* positif sebesar 95%, serta nilai *F-1 Score* untuk *sentiment* negatif sebesar 18% dan untuk *sentiment* positif sebesar 79%. *Recall* dan *F-1 Score* pada *sentiment* negatif masih cukup rendah karena banyak data negatif yang tidak terprediksi secara benar oleh model ini.
2. Model *Naive Bayes*, menghasilkan akurasi sebesar 62%, sehingga tingkat akurasi model RNN *Learning* lebih tinggi.



3. Model LSTM, dengan pengaturan parameter *W2V\_Window* sebesar 6, *W2V\_Epoch* sebesar 32, dan *W2V\_Min\_Count* sebesar 10, *neuron* sebesar 300, dan menggunakan aktivasi *sigmoid* menghasilkan akurasi terbesar bernilai 67%. Kemudian, proses *training* pada pengujian ini menggunakan sebanyak 10 epoch dan 32 batch size. Nilai *precision* yang didapatkan dari model simple RNN untuk *sentiment* negatif sebesar 60% dan untuk *sentiment* positif sebesar 67%. Kemudian untuk nilai *recall* yang didapatkan pada *sentiment* negatif yaitu sebesar 11% dan untuk *sentiment* positif sebesar 96%, serta nilai *F-1 Score* untuk *sentiment* negatif sebesar 18% dan untuk *sentiment* positif sebesar 79%.

Ulasan hasil dari *data testing* divisualisasi menggunakan *wordlink* dan *wordcloud* sesuai masing-masing kelas, diantaranya kelas positif dan kelas negatif. Pada kelas positif, berisi simpulan bahwa masyarakat mendukung kebijakan pemerintah agar tetap menggunakan protokol kesehatan yang mudah diterapkan oleh masyarakat contohnya memakai masker dan cuci tangan untuk mencegah penularan virus corona. Kemudian, pada kelas negatif menghasilkan ulasan yang berisi simpulan bahwa diperlukan kebijakan pemakaian masker dan cuci tangan yang lebih ketat karena isolasi mandiri dianggap sulit diterapkan oleh masyarakat.

## **B. Saran**

Berdasarkan pembahasan dan simpulan yang telah diuraikan, terdapat beberapa hal yang dapat dikembangkan dalam penelitian selanjutnya. Contohnya yaitu dengan menambah jumlah *dataset* untuk menghasilkan akurasi terbaik, serta membangun model dengan metode lain seperti *Support Vector Machine* untuk perbandingan akurasi yang lebih baik.

## DAFTAR PUSTAKA

- Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. 2017. Sentiment Analysis Using Deep Learning Techniques: A Review. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*. 8(6).
- Buslim, N., Busman, B., Sinatrya, N. S., & Kania, T. S. 2018. Analisa Sentimen Menggunakan Data Twitter, Flume, Hive Pada Hadoop dan Java Untuk Deteksi Kemacetan di Jakarta. *Jurnal Online Informatika*, 3(1), 1. <https://doi.org/10.15575/join.v3i1.141>. Diakses pada 22 September 2020.
- Darujati, C. 2016. Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa. *Jurnal Link*, 16, 8.
- Hassan, A. 2017. Sentiment Analysis With Recurrent Neural Network And Unsupervised. *IEEE Signal Processing Society Sigport [Online]*. <https://dx.doi.org/>. Diakses pada 22 September 2020.
- Hong, Z., Wenzhen, J., & Guocai, Y. 2019. An Effective Text Classification Model Based on Ensemble Strategy. *Journal of Physics: Conference Series*, 1229(1). <https://doi.org/10.1088/1742-6596/1229/1/012058>. Diakses pada 5 Mei 2021.
- Jacob, V. E., Lumenta, A. S. M., Jacobus, A., Elektro, T., Sam, U., & Manado, J. K. B. 2019. Rancang Bangun Aplikasi Kemiripan Dokumen Dengan Sumber – Sumber Internet. *Rancang Bangun Aplikasi Kemiripan Dokumen Dengan Sumber – Sumber Internet*, 14(2), 159–164. <https://doi.org/10.35793/jti.14.2.2019.23990>. Diakses pada 22 September 2021.
- Jang, B., Kim, I., & Kim, J. W. 2019. Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS ONE*, 14(8), 1–20. <https://doi.org/10.1371/journal.pone.0220976>. Diakses pada 6 Mei 2021.

- Kementerian Kesehatan RI. 2020. Keputusan Menteri Kesehatan Republik Indonesia Nomor Hk.01.07/Menkes/382/2020 Tentang Protokol Kesehatan Bagi Masyarakat Di Tempat Dan Fasilitas Umum Dalam Rangka Pencegahan Dan Pengendalian Corona Virus Disease 2019 (Covid-19). <https://covid19.kemkes.go.id/>. Diakses pada 22 September 2020.
- Langgeni, D. P., Baizal, Z. K. A., & W, Y. F. A. 2010. Clustering Artikel Berita Berbahasa Indonesia. *Seminar Nasional Informatika, 2010* (semnasIF), 1–10.
- Lestari, M. A., Adikara, P. P., & Adinugroho, S. 2019. *Rekomendasi Lagu berdasarkan Lirik dan Genre Lagu menggunakan Metode Word Embedding (Word2Vec)*. 3(8), 7898–7904.
- Liu, B. 2015. Sentiment analysis: Mining opinions, sentiments, and emotions. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, May*, 1–367. <https://doi.org/10.1017/CBO9781139084789>. Diakses pada 5 Mei 2021.
- Mikolov, Tomas, Chen, K., Corrado, G., & Dean, J. 2013. Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.
- Mikolov, Tomáš, Karafiát, M., Burget, L., Jan, C., & Khudanpur, S. 2010. Recurrent neural network based language model. *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, 1045–1048.
- Monarizqa, N., Nugroho, L. E., & Hantono, B. S. 2014. Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro Dan Teknologi Informasi, 1*, 151–155.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(January), 2825–2830.
- Prasetyo, E. 2012. *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*. ANDI, Jakarta, 360 hlm.

- Ranzato, M. 'A, Hinton, G., & LeCun, Y. 2015. Guest Editorial: Deep Learning. *International Journal of Computer Vision*, 113(1), 1–2. <https://doi.org/10.1007/s11263-015-0813-1>. Diakses pada 5 Mei 2021.
- Risnantoyo, R., Arifin N., Kresna M. 2020. Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm. *JITE (Journal of Informatics and Telecommunication Engineering) Initial Centroid Optimization of K-Means Algorithm Using*. 4(1), 86–96.
- Sari, W. K., Rini, D. P., Malik, R. F., & Azhar, I. S. B. 2021. *JURNAL RESTI Klasifikasi Teks Multilabel pada Artikel Berita Menggunakan Long Short-Term Memory dengan Word2Vec*. 1(10), 276–285.
- Sitohang, V. 2020. Novel Virus Corona (2019-nCoV). *Indo American Journal of Pharmaceutical Sciences*, 23(3), 6. <https://doi.org/10.5281/zenodo.1477753>. Diakses pada 22 Juli 2020.
- Thomas, M., & C.A, L. 2018. Sentimental analysis using recurrent neural network. *International Journal of Engineering & Technology*, 7(2.27), 88. <https://doi.org/10.14419/ijet.v7i2.27.12635>. Diakses pada 22 Juli 2020.
- Twitter, Inc. Authentication. 2020. <http://dev.twitter.com/oauth>. Diakses pada 22 Juli 2020.
- Van, R. G. 2018. Python Tutorial Release 3.7.0 Guido van Rossum and the Python Development Team. [https://bugs.python.org/file47781/Tutorial\\_EDIT.pdf](https://bugs.python.org/file47781/Tutorial_EDIT.pdf). Diakses pada 13 April 2020.