

**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA MASA  
PANDEMI COVID-19 MENGGUNAKAN MEDIA SOSIAL DENGAN  
METODE CONVOLUTIONAL NEURAL NETWORK (CNN)**

**SKRIPSI**

Oleh :  
**LAILA ZAHRONA**  
**NIM. 18650024**



**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**

**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA  
MASA PANDEMI COVID-19 MENGGUNAKAN MEDIA SOSIAL  
DENGAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)**

**SKRIPSI**

Oleh :  
**LAILA ZAHRONA**  
**NIM. 18650024**

**Diajukan kepada:  
Fakultas Sains dan Teknologi UIN Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S. Kom)**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**

## LEMBAR PERSETUJUAN

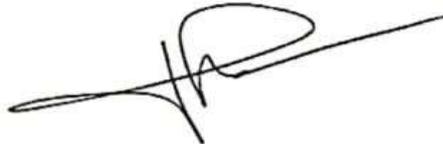
**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA  
MASA PANDEMI COVID-19 MENGGUNAKAN MEDIA SOSIAL  
DENGAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)**

### SKRIPSI

Oleh :  
**LAILA ZAHRONA**  
**NIM. 18650024**

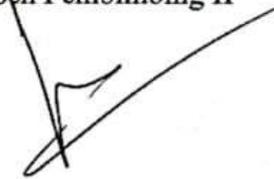
Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal: 15 Juni 2022

Dosen Pembimbing I



**Dr. Fachrul Kurniawan, M.MT**  
NIP. 19771020 200912 1 001

Dosen Pembimbing II



**Irwan Budi Santoso, M.Kom**  
NIP. 19770103 201101 1 004

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
**Dr. Fachrul Kurniawan, M.MT., IPM**  
NIP. 19771020 200912 1 001

## LEMBAR PENGESAHAN

**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA  
MASA PANDEMI COVID-19 MENGGUNAKAN MEDIA SOSIAL  
DENGAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)**

### SKRIPSI

Oleh :

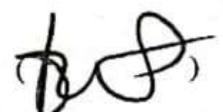
**LAILA ZAHRONA**

**NIM. 18650024**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer (S. Kom)  
Tanggal : 15 Juni 2022

#### Susunan Dewan Penguji

Penguji Utama : Prof. Dr. Suhartono, M.Kom  
NIP. 19680519 200312 1 001  
Ketua Penguji : Fatchurrochman, M.Kom  
NIP. 19700731 200501 1 002  
Sekretaris Penguji : Dr. Fachrul Kurniawan, M.MT  
NIP. 19740510 200501 1 007  
Anggota Penguji : Irwan Budi Santoso, M.Kom  
NIP. 19770103 201101 1 004

()  
()  
()  
()

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT., IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Laila Zahrona

NIM : 18650024

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Judul Skripsi : Klasifikasi Kalimat Perbincangan Masyarakat Pada Masa Pandemi Covid-19 Menggunakan Media Sosial Dengan Metode Convolutional Neural Network (CNN)

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar – benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 Juni 2022

Yang membuat pernyataan,



Laila Zahrona  
NIM. 18650024

## **MOTTO**

*“ Sing Penting Yakin ”*

## **HALAMAN PERSEMBAHAN**

Puji syukur kehadiran Allah SWT, atas limpahan karunia-Nya saya dapat menyelesaikan skripsi ini. Shalawat serta salam tak lupa saya ucapkan kepada Nabi Muhammad SAW. Saya persembahkan skripsi ini kepada kedua orang tua saya yang menjadi motivasi utama untuk mengerjakan skripsi ini. Kepada Ayah Muhaimin dan Ibu Erni Wahyu Lindati yang tiada henti selalu mendoakan dan mendukung saya hingga tahap ini dan tahap selanjutnya, saya mengucapkan terima kasih yang sangat mendalam. Kepada keluarga besar yang mendoakan dan membantu saya dalam penyelesaian skripsi ini. Semoga kebaikan, rahmat dan perlindungan Allah SWT senantiasa tercurah kepada mereka semua. Amin.

## KATA PENGANTAR

*Assalamualaikum Warahmatullahi Wabarakatuh.*

Syukur alhamdulillah penulis haturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Kalimat Perbincangan Masyarakat pada Masa Pandemi Covid-19 Menggunakan Media Sosial dengan Metode *Convolutional Neural Network* (CNN)” sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik Informatika jenjang Strata-1 Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penulis menyadari bahwa dalam penyusunan skripsi ini tak luput dari dukungan dan bantuan dari berbagai pihak. Oleh karena itu, dengan segenap kerendahan hati penulis mengucapkan terima kasih kepada:

1. Kedua orang tua penulis yang telah memberikan dukungan baik secara moril maupun materil sehingga penulisan dapat menyelesaikan skripsi ini.
2. Prof. Dr. H. M. Zainuddin, MA, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang beserta jajarannya.
3. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang beserta jajarannya.
4. Dr. Fachrul Kurniawan, M.MT, IPM selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang dan sekaligus sebagai Dosen Pembimbing 1, yang telah meluangkan waktu untuk membimbing dan memberikan masukan serta arahan selama penyusunan skripsi hingga selesai.
5. Irwan Budi Santoso, M.Kom selaku Dosen Pembimbing 2, yang telah meluangkan waktu untuk membimbing dan memberikan masukan selama penyusunan skripsi hingga selesai.
6. Prof. Dr. Suhartono, M.Kom dan Fatchurrochman, M.Kom selaku Dosen Penguji yang telah memberikan saran dan arahan kepada penulis dalam proses penyelesaian skripsi ini.
7. Seluruh Dosen dan Jajaran Staf Jurusan Teknik Informatika yang telah memberikan ilmu dan informasi yang sangat bermanfaat kepada saya, serta secara tidak langsung ikut terlibat dalam penyusunan skripsi ini.
8. Teman-teman dekat saya Febrina Dwita Sari, Wahyu Febriansyah, Dwi Zulva Ulinuha, Yuliana Romadhoni, Nurul Hidayah, Uswa, Nadia, Faida

dan teman-teman lain yang selalu memberi semangat untuk menyelesaikan tugas akhir ini.

9. Saudara Teknik Informatika Angkatan 2018 “UFO” yang telah membantu bertukar pikiran dan mendukung dalam proses penyelesaian skripsi.
10. Tante Ratna dan Om Catur yang memberikan tempat tinggal kepada penulis dan teman-teman penulis selama di Kota Malang
11. Tak lupa teman – teman terdekat penulis yang selalu mendoakan dan memberikan motivasi yang dapat membangun semangat penulis dalam menyelesaikan skripsi ini.
12. Semua pihak yang ikut membantu dalam penyusunan skripsi hingga selesai.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan karena terdapat peribahasa “tak ada gading yang tak retak” yang memiliki arti bahwa tidak ada sesuatu yang tidak memiliki cacat dan tidak ada sesuatu yang sempurna. Oleh karena itu, penulis berharap semoga skripsi ini dapat memberikan manfaat kepada para pembaca termasuk penulis sendiri.

*Wassalamualaikum Warahmatullahi Wabarakatuh.*

## DAFTAR ISI

HALAMAN JUDUL .....	iii
LEMBAR PERSETUJUAN .....	iii
LEMBAR PENGESAHAN .....	iii
MOTTO .....	v
PERNYATAAN KEASLIAN TULISAN .....	vi
HALAMAN PERSEMBAHAN .....	vi
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL .....	xii
ABSTRAK .....	xiii
ABSTRACT .....	xiv
مستخلص البحث.....	xv
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Tujuan Penelitian .....	7
1.4 Batasan Masalah .....	7
1.5 Manfaat Penelitian .....	7
<b>BAB II STUDI PUSTAKA .....</b>	<b>8</b>
2.1 Penelitian Terkait .....	8
2.2 Landasan Teori.....	10
2.2.1 <i>Deep Learning</i> .....	10
2.2.2 <i>Natural Language Processing (NLP)</i> .....	14
2.2.3 <i>Preprocessing</i> .....	15
2.2.4 <i>Convolutional Neural Network (CNN)</i> .....	16
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>22</b>
3.1 Tahapan Penelitian .....	22
3.2 Perancangan Sistem .....	24
3.3 <i>Collecting Data</i> .....	25
3.4 <i>Preprocessing</i> .....	26
3.4.1 <i>Cleaning</i> .....	27
3.4.2 <i>Case Folding</i> .....	27
3.4.3 <i>Tokenizing</i> .....	28
3.4.4 <i>Stemming</i> .....	28
3.5 <i>Word Embedding</i> .....	29
3.6 Proses Perhitungan Metode <i>Convolutional Neural Network</i> .....	30
3.7 Skenario Pengujian .....	34
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>	<b>36</b>
4.1 Skenario Uji Coba .....	36
4.1.1 <i>Collecting Data</i> .....	36
4.1.2 <i>Preprocessing</i> .....	37
4.1.3 <i>Split Data</i> .....	39

4.1.4	Membangun Model Convolutional Neural Network .....	42
4.2	Hasil Uji Coba.....	46
4.3	Pembahasan.....	49
<b>BAB V</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>53</b>
5.1	Kesimpulan .....	53
5.2	Saran .....	54
<b>DAFTAR PUSTAKA</b>		

## DAFTAR GAMBAR

Gambar 2.1 Model <i>Deep Convolutional Neural Network</i> .....	18
Gambar 2.2 Struktur Pada ReLU .....	18
Gambar 2.3 Struktur Pada Tanh.....	20
Gambar 3.1 Tahapan Penelitian .....	22
Gambar 3.2 Desain Sistem.....	24
Gambar 3.3 Proses Perhitungan Konvolusi .....	31
Gambar 3.4 Proses <i>Activation Function</i> (ReLU).....	32
Gambar 3.5 Proses <i>Pooling Layer</i> .....	32
Gambar 3.6 Contoh <i>fully connected layer</i> .....	33
Gambar 4.1 <i>Library Tweepy</i> .....	36
Gambar 4.2 API twitter .....	36
Gambar 4.3 Kode program <i>cleaning</i> .....	37
Gambar 4.4 Kode program <i>case folding</i> .....	38
Gambar 4.5 Kode program <i>tokenizing</i> .....	38
Gambar 4.6 Kode program <i>stemming</i> .....	38
Gambar 4.7 <i>Text Sequences</i> .....	41
Gambar 4.8 <i>Padding</i> .....	41
Gambar 4.9 <i>Split Data</i> .....	42
Gambar 4.10 Arsitektur model CNN .....	43
Gambar 4.11 Kode program model CNN .....	44
Gambar 4.12 Kode program pelatihan model CNN.....	44
Gambar 4.13 Grafik Loss dan Akurasi .....	46
Gambar 4.14 Hasil <i>confusion matrix</i> .....	47
Gambar 4.15 Nilai Evaluasi .....	47

## DAFTAR TABEL

Tabel 2.1 Formula Evaluasi Kinerja Klasifikasi .....	12
Tabel 2.2 <i>Confusion Matriks</i> .....	13
Tabel 3.1 Data Terlabel .....	25
Tabel 3.2 <i>Cleaning</i> .....	26
Tabel 3.3 <i>Case Folding</i> .....	28
Tabel 3.4 <i>Tokenizing</i> .....	28
Tabel 3.5 <i>Stemming</i> .....	29
Tabel 3.6 Kalimat Negatif .....	29
Tabel 3.7 Vektor Kata .....	30
Tabel 3.8 <i>Confusion Matrix</i> .....	34
Tabel 4.1 Jumlah Data Terlabel .....	39
Tabel 4.2 Indek kata .....	42
Tabel 4.3 Hasil <i>Training Split Data</i> Model A .....	288
Tabel 4.4 Hasil <i>Training Split Data</i> Model B .....	28
Tabel 4.5 Klasifikasi pada Sistem .....	28

## ABSTRAK

Zahrona, Laila. 2022. *Klasifikasi Kalimat Perbincangan Masyarakat pada Masa Pandemi Covid-19 Menggunakan Media Sosial dengan Metode Convolutional Neural Network (CNN)*. Skripsi. Jurusan Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fachrul Kurniawan, M.MT (II) Irwan Budi Santoso, M.Kom.

---

**Kata Kunci:** *Klasifikasi, Media Sosial, Convolutional Neural Network (CNN)*.

Media sosial merupakan sebuah sarana online yang digunakan untuk interaksi oleh semua lapisan masyarakat secara bebas dimanapun dan kapanpun, pada sisi lain media sosial dapat digunakan untuk membagikan dan menyebarluaskan informasi hingga berita. Media sosial yang pada saat ini sering digunakan oleh pengguna yaitu twitter. Media sosial twitter itu sendiri dapat menemukan apa yang sedang terjadi secara langsung dan menemukan sumber informasi tepercaya. Informasi yang sedang ramai diperbincangkan yaitu tentang pandemi covid-19. Adanya pandemi covid-19, seluruh aturan dan larangan banyak diterapkan di Negara-Negara yang terdampak salah satunya di negara kita yaitu Indonesia. Maka dari itu banyaknya masyarakat yang menyampaikan tanggapan seputar covid-19 melalui media sosial dibuatlah penelitian klasifikasi kalimat perbincangan masyarakat terhadap topik yang sedang ramai dibicarakan yaitu tentang pandemi covid-19 dengan algoritma *Convolutional Neural Network*. Dengan dilakukannya penelitian ini diharapkan dapat digunakan oleh lembaga dan instansi pemerintah menilai opini masyarakat terhadap pandemic covid secara lebih benar untuk evaluasi pelayanan terhadap pandemi masyarakat. Pada penelitian ini membahas algoritma *Convolutional Neural Network* yang digunakan sebagai klasifikasi kalimat perbincangan masyarakat tentang pandemi covid-19 dengan media sosial yang dipakai adalah twitter. Input data pada penelitian ini adalah data dari media sosial yang di proses menggunakan metode *preprocessing* dan terdapat label klasifikasi yang diolah oleh pakar pada bidang bahasa, sedangkan outputnya adalah hasil akurasi dari pelatihan dan pengujian algoritma *Convolutional Neural Network*. Pada proses pelatihan dilakukan hingga menghasilkan arsitektur jaringan yang tepat, kemudian untuk proses pengujian dilakukan untuk melihat performa akurasi *Convolutional Neural Network* dalam klasifikasi kalimat perbincangan pada masa pandemi covid-19. Pada proses *training* data terdapat 567 data yang digunakan sebagai data uji, dari data tersebut menghasilkan akurasi sebesar 81% dalam menentukan proses klasifikasi terkait pandemi covid. Selain itu, pada penelitian ini juga melakukan percobaan klasifikasi dari 15 *tweet* oleh sistem didapatkan hasil yang sama dengan data yang diklasifikasi manual oleh pakar dalam bidang bahasa.

## ABSTRAK

Zahrona, Laila. 2022. *Klasifikasi Kalimat Perbincangan Masyarakat pada Masa Pandemi Covid-19 Menggunakan Media Sosial dengan Metode Convolutional Neural Network (CNN)*. Skripsi. Jurusan Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fachrul Kurniawan, M.MT (II) Irwan Budi Santoso, M.Kom.

---

**Kata Kunci:** *Klasifikasi, Media Sosial, Convolutional Neural Network (CNN)*.

Media sosial merupakan sebuah sarana online yang digunakan untuk interaksi oleh semua lapisan masyarakat secara bebas dimanapun dan kapanpun, pada sisi lain media sosial dapat digunakan untuk membagikan dan menyebarluaskan informasi hingga berita. Media sosial yang pada saat ini sering digunakan oleh pengguna yaitu twitter. Media sosial twitter itu sendiri dapat menemukan apa yang sedang terjadi secara langsung dan menemukan sumber informasi tepercaya. Informasi yang sedang ramai diperbincangkan yaitu tentang pandemi covid-19. Adanya pandemi covid-19, seluruh aturan dan larangan banyak diterapkan di Negara-Negara yang terdampak salah satunya di negara kita yaitu Indonesia. Maka dari itu banyak masyarakat yang menyampaikan tanggapan seputar covid-19 melalui media sosial dibuatlah penelitian klasifikasi kalimat perbincangan masyarakat terhadap topik yang sedang ramai dibicarakan yaitu tentang pandemi covid-19 dengan algoritma *Convolutional Neural Network*. Dengan dilakukannya penelitian ini diharapkan dapat digunakan oleh lembaga dan instansi pemerintah menilai opini masyarakat terhadap pandemic covid secara lebih benar untuk evaluasi pelayanan terhadap pandemi masyarakat. Pada penelitian ini membahas algoritma *Convolutional Neural Network* yang digunakan sebagai klasifikasi kalimat perbincangan masyarakat tentang pandemi covid-19 dengan media sosial yang dipakai adalah twitter. Input data pada penelitian ini adalah data dari media sosial yang di proses menggunakan metode *preprocessing* dan terdapat label klasifikasi yang diolah oleh pakar pada bidang bahasa, sedangkan outputnya adalah hasil akurasi dari pelatihan dan pengujian algoritma *Convolutional Neural Network*. Pada proses pelatihan dilakukan hingga menghasilkan arsitektur jaringan yang tepat, kemudian untuk proses pengujian dilakukan untuk melihat performa akurasi *Convolutional Neural Network* dalam klasifikasi kalimat perbincangan pada masa pandemi covid-19. Pada proses *training* data terdapat 567 data yang digunakan sebagai data uji, dari data tersebut menghasilkan akurasi sebesar 81% dalam menentukan proses klasifikasi terkait pandemi covid. Selain itu, pada penelitian ini juga melakukan percobaan klasifikasi dari 15 *tweet* oleh sistem didapatkan hasil yang sama dengan data yang diklasifikasi manual oleh pakar dalam bidang bahasa.

## ABSTRACT

Zahrona, Laila. 2022. *Classification of Public Conversation Sentences during the Pandemic Covid-19 Using Social Media with the Convolutional Neural Network (CNN) Algorithm*. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Counselor : (I) Dr. Fachrul Kurniawan, M.MT (II) Irwan Budi Santoso, M.Kom.

---

**Keywords** : *Classification, Social Media, Convolutional Neural Network (CNN)*.

Social media is an online tool that is used for interaction by all levels of society freely wherever and whenever, social media also can be accustomed share and pass around information. One of the most widely use the social media is twitter. Twitter may be a social media wherever you will be able to determine what is going on in real time and find trusted sources of information. The information that is currently being discussed is the corona virus disease. Due to the Covid-19 pandemic, all rules and restrictions have been implemented in many affected countries, one of the country is Indonesia. Therefore, there are many people who provide feedback about covid-19 through social media, this research is made on the classification of community conversation sentences on the topic that is being discussed, namely the Covid-19 pandemic with the Convolutional Neural Network algorithm. With this research, it is hoped that it can be used by government institutions and agencies to assess public opinion on the object of a pandemic more precisely for evaluating the government services for pandemics. In this study, we discuss the Convolutional Neural Network algorithm that is used to classify sentences in public conversations about the COVID-19 pandemic with twitter as the social media. The input data in this study is data from social media which is processed using the preprocessing method and there is a classification label that is processed by experts in the language field, while the output is the result of accuracy from training and testing Convolutional Neural Network or CNN algorithm. The training process is carried out to produce the right network architecture, then the testing process is carried out to see the accuracy performance of the Convolutional Neural Network in the classification of conversational sentences during the covid-19 pandemic. In the testing process there are 567 data used as test data, from this data it produces an accuracy of 81% in determining the classification process related to the covid-19 pandemic. Moreover, this study also conducted an experiment on the classification of 15 tweets by the system, and the results were the same as those classified manually by experts in the field of language.

## مستخلص البحث

زهرنا, ليلي. ٢٠٢٢. تصنيف جمل المناقشة العامة خلال وباء كوفيد ١٩ باستخدام وسائل التواصل الاجتماعي بالطريقة الشبكة العصبية التلافيفية . المناقشة . قسم هندسة المعلومات. كلية العلم تكنولوجيا . الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مانج المشريف : (١) الدكتور فحل كورنياوان , الماجستير. المشريف (٢) إروان بودي سنتوسو الماجستير .

**الكلمات الرئيسية:** تصنيف, وسائل التواصل الاجتماعي, الشبكة العصبية التلافيفية

وسائل التواصل الاجتماعي هي وسيلة على الإنترنت تستخدم للتفاعل الحر من قبل جميع مستويات المجتمع في كل مكان وفي أي وقت، ويمكن أيضا استخدام وسائل التواصل الاجتماعي لتوزيع ونشر المعلومات. وسائل التواصل الاجتماعي حاليا على نطاق واسع، من بين أمور أخرى، تويتر. تويتر هو وسائل التواصل الاجتماعي التي يمكنها معرفة ما يحدث بشكل مباشر وإيجاد معلومات موثوقة. وكانت المعلومات المثيرة للجدل هي وباء كوفيد ١٩. تم تنفيذ جائحة كوفيد ١٩، المجموعة الكاملة من القوانين والمحظورات، في دول تأثرت بواحد إندونيسي. بناء على ذلك، يتم إجراء العديد من المجتمعات التي تعالج كوفيد ١٩ عبر وسائل التواصل الاجتماعي من خلال دراسات لتصنيف جمل المناقشة العامة حول الموضوع قيد المناقشة حول خوارزمية كوفيد ١٩ مع خوارزمية بالطريقة الشبكة العصبية التلافيفية على أمل أن تقوم وكالاتها والوكالات الحكومية بالحكم على الرأي العام من موضوع الوباء بشكل أكثر دقة لتقييم خدمات مجتمع الوباء. في هذه الدراسة تناقش الشبكة العصبية التلافيفية التابع لها لتصنيف جمل محادثة المجتمع بشأن الجائحة كوفيد ١٩ مع وسائل الاعلام الاجتماعية المستخدمة تويتر. بيانات الإدخال في هذه الدراسة هي بيانات من وسائل الاعلام الاجتماعية في عملية استخدام طريقة بروجروسينغ وهناك تصنيف تصنيف معالجتها من قبل الخبراء في لغة اللغة، في حين أن الناتج هو نتيجة دقة التدريب واختبار الشبكة العصبية التلافيفية غير قانوني. في عملية التدريب يتم حتى تنتج الهندسة المعمارية للشبكة اليمنى، ثم لعملية الاختبار تتمثل في رؤية أداء دقة الشبكة العصبية التلافيفية في تصنيف أحكام المحادثة في الوكالة الكوفية ١٩ في عملية الاختبار هناك ٥٦٧ بيانات تستخدم كبيرة الاختبار، من البيانات الناتجة عن دقة ٨١% في تحديد عملية تصنيف الوباء الصريح المرتبط ١٩. وبالإضافة إلى ذلك، أجريت في هذه الدراسة أيضا تجربة التصنيف من ١٥ تغريدات من قبل النظام حصل على نفس النتيجة كما البيانات المصنفة من قبل الخبراء في اللغة.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pandemi covid-19 atau yang dikenal wabah virus corona saat ini menjadi permasalahan global. Permasalahan pandemi virus corona yang saat ini masih belum selesai dan dampaknya yang begitu luar biasa bukan hanya pada kesehatan masyarakat tapi berdampak juga pada semua bidang kehidupan masyarakat pada umumnya baik bidang politik, ekonomi, sosial, dan budaya. Dengan banyaknya korban jiwa bahkan merenggut banyak nyawa dalam waktu singkat, pandemi ini menjadi terror mengerikan bagi semua masyarakat di seluruh dunia. Virus ini awal mulanya terjadi pada tanggal 31 Desember 2019, otoritas Tiongkok melaporkan kasus pneumonia di Wuhan, Cina. Sebagian besar pasien yang melapor terpapar dari pasar makanan laut besar yang menjual banyak spesies hewan hidup. Berdasarkan pengalaman dari wabah *CoV zoonosis* sebelumnya, otoritas kesehatan masyarakat telah memulai kegiatan kesiapsiagaan dan respon. Para pemimpin Wuhan menutup dan mendisinfektan pasar pertama yang teridentifikasi. Amerika Serikat dan beberapa negara lain telah memulai penyaringan masuk penumpang dari Wuhan di pelabuhan masuk utama. Praktisi kesehatan di kota-kota China lainnya, Thailand, Jepang, dan Korea Selatan segera mengidentifikasi kasus terkait perjalanan, mengisolasi individu untuk perawatan lebih lanjut. Kasus terkait perjalanan pertama di Amerika Serikat terjadi pada 21 Januari pada seorang pria muda China yang telah mengunjungi Wuhan (Paules et al., 2020).

Disamping itu, ketika virus corona melanda China sejak bulan Desember

2019 - Februari 2020, Indonesia melaporkan bahwa tidak ada masyarakat yang terinfeksi sama sekali. Akhirnya pada tanggal 2 Maret 2020, Kepala negara Indonesia, Ir. Joko Widodo mengkonfirmasi dua orang yang terjangkit covid pertama di Indonesia. Hingga pada tanggal 27 Maret 2020, negara Indonesia telah mencapai 103 kasus baru, 893 kasus yang dikonfirmasi, dengan 78 jumlah kematian, dan 35 sembuh. Sebagai upaya menekan siklus penyebaran virus corona, beberapa pemerintah daerah memberlakukan aturan baru untuk kegiatan belajar dilakukan dari rumah bagi seluruh siswa sekolah. Aktivitas belajar mengajar di seluruh tingkat pendidikan diliburkan selama jangka waktu 14 hari. Begitu pula para pekerja. Seluruh perusahaan juga membuat kebijakan untuk seluruh karyawannya mengerjakan pekerjaan mereka dari rumah atau yang biasa disebut *work from home* selama rentang waktu 14 hari. Banyak instansi yang memberlakukan kebijakan tersebut sejak tanggal 16 Maret 2020, ketika eskalasi virus corona di beberapa daerah mengalami kenaikan. Selain himbauan untuk melakukan segala aktivitas di rumah, masyarakat juga diwajibkan disiplin untuk melakukan jaga jarak atau yang biasa disebut *sosial distancing* berjarak satu meter saat sedang berada dikeramaian atau di suatu tempat bersama orang lain, terutama pada ruang publik (Kemenkes, 2020).

Dengan adanya kasus covid-19 yang masih mewabah walaupun pemerintah sudah membuat kebijakan untuk menekan kasus agar lebih membaik dan vaksin juga sudah ditemukan namun berbagai varian baru dari virus corona seperti *Alpha*, *Beta*, dan *Delta* ini bermunculan membuat banyak masyarakat yang terjangkit virus baru tersebut dan terjadi lonjakan korban jiwa di setiap daerah. Disamping pandemi

terus menerus menyebabkan korban jiwa yang merenggut banyak nyawa, pemerintah juga bergerak cepat membuat kebijakan baru seperti kebijakan Kebijakan pemberlakuan pembatasan kegiatan masyarakat. Kebijakan pemberlakuan pembatasan kegiatan masyarakat atau yang disingkat PPKM ini disesuaikan oleh level pada situasi di masing-masing daerah. Selain itu juga diukur oleh indikator konfirmasi infeksi virus perhari, jumlah terisinya ruangan di rumah sakit atau *Bed Occupancy Ratio* (BOR), dan pencapaian penekanan vaksinasi (Kemenkes, 2021). Walaupun pemerintah mengeluarkan kebijakan untuk menekan agar pandemi ini lekas membaik akan tetapi tidak sedikit yang beropini bahkan menyebarkan berita palsu bahwa kasus covid-19 ini hanyalah akal-akalan pemerintah, pebisnis hingga tim medis untuk mendapatkan keuntungan dari kasus ini. Padahal dalam agama islam diajarkan bahwa sebagai manusia harus menyebarkan berita yang benar dan sebelum menerima berita harus meneliti apakah berita itu benar atau tidak. Pada QS Al-Hujurat Ayat 6 Allah Swt. Berfirman :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحِحُوا عَلَىٰ مَا فَعَلْتُمْ نُدْمِينَ

*“Wahai orang-orang yang beriman! Jika terdapat seseorang yang fasik datang kepadamu membawa berita, maka telitilah kebenaran berita tersebut, agar kamu tidak mencelakakan suatu kaum karena kecerobohan, yang akhirnya kamu menyesali perbuatanmu itu”* (Qs. al- Hujurat /49:6).

Dalam tafsir Jalalain, ayat diatas memiliki tafsir sebagai berikut “wahai orang-orang mukmin! Apabila datang kepadamu orang yang fasik membawa berita, maka telitilah kebenaran beritanya tersebut oleh kalian, apakah mereka yang membawa beria benar atau dusta. Menurut dari suatu sumber, apabila dibaca Fatatsabbatuu yang berasal dari lafal Ats-Tsabaat yang memiliki arti yaitu teliti terlebih dahulu untuk kebenarannya. Hal tersebut bermaksud supaya kalian tidak memberikan

musibah pada suatu kaum) menjadi Maf'ul dari lafal Fatabayyanuu, yakni khawatir akan hal tersebut membawa dampak musibah kepada suatu kaum dengan tidak mengetahui bagaimana keadaan mereka dan akan menjadi Fail, yaitu tanpa sepengetahuannya yang membuat seseorang telah berbuat kesalahan pada suatu kaum.

Dalam Tafsir Al-Misbah menjelaskan bahwa Q.S Al-Hujurat ayat 6 bahwa sebelum menerima informasi harus diteliti terlebih dahulu, informasi atau berita yang didapatkan tidak harus langsung diterima begitu saja dan harus diteliti kebenarannya, apalagi jika yang menyebarkan berita itu adalah orang fasik, maka harus diteliti lagi untuk kebenaran, apakah di kehidupan sehari-hari penyebar berita sering berbohong atau tidak dan dapat dipercaya atau tidak, karena belum tentu yang diucapkan benar berasal dari orang lain yang tidak mengetahui berita tersebut secara pasti. Karena bisa jadi penyebar berita tidak mengerti masalah atau persoalan tersebut, bisa jadi mereka telah memiliki asumsi dan pemikiran yang keliru.

Dengan ayat tersebut kita dapat menyimpulkan bahwa kita harus meneliti kembali suatu berita yang disebar. Kita harus melihat berita tersebut benar atau tidak apalagi pada waktu pandemi ini berita palsu sangat banyak beredar di *platform* media sosial karena berita palsu tersebut bisa saja merusak kehidupan bermasyarakat. Media sosial saat ini berkembang sangat pesat mengingat media sosial merupakan salah satu perkembangan teknologi komunikasi dan informasi. Media sosial menjadi alat komunikasi dan membagikan informasi yang efektif digunakan, terbukti dahulu setiap orang yang melakukan komunikasi harus menggunakan surat. Namun, saat ini seseorang yang memiliki media sosial dapat

melakukan komunikasi dan menyebarkan informasi setiap saat. Kemudahan tersebut menjadikan pengguna media sosial semakin meningkat setiap tahunnya. Media sosial bukan hanya dijadikan untuk media komunikasi. namun, juga dijadikan untuk tempat penyebaran informasi. Salah satu *platform* media sosial yang penyebaran informasinya terbukti cepat adalah twitter. Twitter adalah salah satu platform global untuk ekspresi diri dan percakapan publik secara *real time*. Twitter mengizinkan orang untuk membuat, melihat, membagikan dan mendapatkan konten dan telah mendemokratisasikan pembuatan dan distribusi konten. Melalui topik, minat, dan tren, twitter membantu orang menemukan apa yang sedang terjadi secara langsung. Pada tahun 2020, twitter membuat suatu progress baru untuk melayani percakapan publik dengan membantu orang menemukan sumber informasi terpercaya, mengatur dan menampilkan lebih baik banyak topik dan minat yang membawa orang ke Twitter (Annual Report Twitter, 2020). Melihat bahwa twitter merupakan media sosial yang dapat menemukan apa yang sedang terjadi secara langsung dan menemukan sumber informasi terpercaya maka media sosial twitter dapat dijadikan sebagai sumber untuk klasifikasi perbincangan masyarakat terhadap topik yang sedang ramai dibicarakan yaitu tentang pandemi covid-19.

Dari permasalahan diatas pada penelitian ini dibuat suatu solusi dalam melakukan klasifikasi kalimat perbincangan masyarakat terhadap topik yang sedang ramai dibicarakan yaitu tentang pandemi covid-19. Penelitian ini menggunakan metode klasifikasi teks *deep learning* yang merupakan pengembangan dari *machine learning*.

Istilah *Deep Learning* atau *Deep Neural Network* mengacu pada *Artificial Neural Networks* (ANN) dengan *multilayer*. Selama beberapa dekade terakhir, *Deep Learning* telah dianggap sebagai metode yang bagus, dan telah menjadi populer dalam literatur karena kemampuannya untuk memproses data dalam jumlah besar. Salah satu *Deep learning* yang paling populer adalah *Convolutional Neural Network* (CNN). Kata *Convolutional Neural Network* (CNN) berasal dari operasi linier matematis antara suatu matrik yang disebut konvolusi. Metode CNN memiliki banyak lapisan; diantaranya lapisan *convolutional*, lapisan *non-linear*, lapisan *pooling* dan lapisan *fullyconnected*. Lapisan *convolutional* dan *fullyconnected* mempunyai suatu parameter tetapi lapisan *pooling* dan *non-linearitas* tidak punya parameter. Metode CNN memiliki kinerja yang sangat baik dalam masalah pembelajaran mesin (Albawi et al., 2018).

Dalam penelitian ini untuk mendapatkan hasil *accuracy* pada algoritma teks *classification*, algoritma yang digunakan ialah CNN. Diharapkan dari klasifikasi kalimat perbincangan masyarakat terhadap topik yang sedang ramai dibicarakan yaitu tentang pandemi covid-19 dengan algoritma *Convolutional Neural Network* dengan harapan dapat digunakan oleh lembaga kesehatan dan instansi pemerintah menilai opini pada publik terhadap pandemi secara lebih benar untuk evaluasi pelayanan terhadap pandemi masyarakat.

## **1.2 Rumusan Masalah**

- 1) Bagaimana metode yang digunakan untuk mengolah kumpulan data dari twitter sesuai topik yang ditentukan ?
- 2) Bagaimana hasil implementasi *Convolutional Neural Network* dalam

klasifikasi kalimat perbincangan pada masa pandemi covid-19 ?

### **1.3 Tujuan Penelitian**

- 1) Menerapkan metode yang digunakan untuk mengolah kumpulan data dari twitter sesuai topik yang ditentukan.
- 2) Mengetahui performa akurasi *Convolutional Neural Network* dalam klasifikasi kalimat perbincangan pada masa pandemi covid-19.

### **1.4 Batasan Masalah**

- 1) Kumpulan data yang diperoleh dari twitter berbahasa Indonesia sebanyak 2.835 *tweet*.
- 2) Topik yang digunakan pada penelitian ini terkait dengan pandemi covid-19 sejak bulan September 2020 hingga September 2021.
- 3) Algoritma yang digunakan dalam proses klasifikasi adalah *Convolutional Neural Network*.

### **1.5 Manfaat Penelitian**

- 1) Untuk mengetahui metode *deep learning* khususnya pada klasifikasi kalimat menggunakan metode *Convolutional Neural Network*.
- 2) Mendapatkan hasil klasifikasi kalimat perbincangan pada masa pandemi covid-19 menggunakan media sosial twitter.

## **BAB II**

### **TINJAUAN PUSTAKA**

Pada bab ini berisi tinjauan pustaka yang membahas beberapa penelitian terkait, landasan teori, dan metode yang digunakan penulis sebagai acuan mengerjakan penelitian tugas akhir ini.

#### **2.1 Penelitian Terkait**

Terdapat penelitian terdahulu yang terkait dengan penelitian ini mengusulkan menggunakan metode klasifikasi kalimat ilmiah menggunakan satu dimensi *Convolutional Neural Networks*. Data penelitian tersebut diambil dari *sentence* ilmiah yang menggunakan bahasa Inggris dari penelitian mengenai komputasi yang tersedia pada *corpus*. Tahapan pemrosesan data yang digunakan terdiri dari *text cleaning*, *case folding*, dan *tokenizing*. Pada penelitian tersebut mengembangkan *One Dimensions Convolutional Neural Networks* dengan model *Multichannel* dan *Singlechannel* dengan ditambahkan teknik *embedding* menggunakan *Word2vec*. Hasil yang didapatkan pada penelitian tersebut menunjukkan klasifikasi dengan *CNNMultichannel* menghasilkan akurasi 74,51% namun untuk *CNN-Singlechannel* lebih rendah yaitu akurasi sebesar 70,76%. Dalam tahapan pengujian yang telah dilakukan didapatkan konfigurasi terbaik oleh model *CNN-Multichannel* dan menggunakan metode *adam* sebagai optimasi, *learning-rate* sebesar 0,001, *batch size* sebesar 64, dan ukuran *vector* sebesar 100 (Mahendra et al., 2020). Selain dari penelitian diatas, pada penelitian lain juga melakukan *classification* terhadap kalimat retorik dengan metode yang digunakan adalah SVM dan metode *word embedding* adalah *word2vec* dan *glove*. Yang mendapatkan hasil akurasi tertinggi

pada pengujian yang dilakukan didapatkan oleh *word2vec* Model BOW dengan *F1-Score* sebesar 43,44% dengan menggunakan dataset sebesar 20.155 dari penelitian ilmiah (Ranchman et al., 2018).

Penelitian yang dilakukan oleh (Juwiantho et al., 2020) yang membahas tentang *sentiment analysis* berbahasa indonesia dengan *word2vec* dan memakai algoritma *Convolutional Neural Network* pada twitter terhadap opini dari pelanggan bertujuan agar instansi dapat langsung mengetahui tentang kepuasan pada pelanggan dan dipakai untuk melakukan perbaikan pelayanan hingga menaikkan merk perusahaan. Pada penelitian tersebut menggunakan metode *Convolutional Neural Network* karena penggunaan metode *machine learning* klasik banyak yang menerapkannya pada analisis sentimen, namun metode ini tidak terlalu memperhitungkan pentingnya urutan kata pada kalimat dan juga menggunakan model *word2vec* untuk bahasa Indonesia sebagai representasi kata dalam bentuk vektor. Pada penelitian tersebut menggunakan 999 *tweet* Bahasa Indonesia yang diambil dari media sosial Twitter. Hasil yang dilakukan pada penelitian tersebut dengan algoritma *Convolutional Neural Network* (CNN) menghasilkan suatu nilai *accuracy* terbaik sebesar 76,40%.

Penelitian terhadap Sentimen pada pandemi covid-19 dengan *Convolution Neural Network* (CNN) dengan Teknik NLP yang bertujuan untuk mendapatkan persepsi atau gambaran dari masyarakat terkait pilkada pada tahun 2020 di masa pandemi covid-19 menggunakan data dari twitter dengan cara crawling mendapatkan 500 *dataset* twitter. Data tersebut dilabeli dan dikelompokkan dalam 2 kategori sentimen yaitu kelas positif dan negatif yang selanjutnya dilakukan tahap

*preprocessing*. Hasil dari pengujian didapatkan *accuracy* tertinggi sebesar 90% dengan konvolusi sebanyak 4 lapisan dan 100 *epoch*. Hasil pengujian *classification* pada kalimat baru didapatkan prediksi yang sesuai dengan label aslinya (Listyarini & Anggoro, 2021).

## **2.2 Landasan Teori**

Landasan teori membahas beberapa teori yang dipakai dalam pembuatan penelitian tugas akhir ini. Pada landasan teori, penulis menggunakan penelitian dan buku sebagai sumber dan acuan.

### **2.2.1 Deep Learning**

*Deep learning* atau yang dikenal sebagai *deep structured learning* yang merupakan subbidang dari pembelajaran mesin dimana terinspirasi dari fungsi otak manusia. Sama seperti *neuron* yang saling berhubungan di otak. Setiap *neuron* mengambil *input*, melakukan beberapa manipulasi dalam neuron, dan menghasilkan output yang lebih dekat dengan output yang diharapkan dalam hal data yang berlabel (Kulkarni & Shivananda, 2021).

*Deep Learning* atau pembelajaran mendalam merupakan salah satu dari *machine learning* yang berisi dari banyak lapisan dan membentuk suatu tumpukan. Dalam pembelajaran mesin terdapat teknik yang digunakan untuk mengekstraksi fitur dari data *training* dan algoritma pembelajaran khusus untuk membuat klasifikasi dan juga untuk mengetahui suara. Akan tetapi, pada metode tersebut terdapat beberapa kekurangan baik dalam hal kecepatan waktu dan *accuracy*. Sistem konsep JST yang dapat

ditanggungkan pada algoritma *machine learning* yang sudah tersedia sehingga *gadget* saat ini dapat belajar dengan kecepatan waktu, *accuracy*, dan skala yang lebih besar. Prinsip ini berkembang sampai *deep learning* sering dipakai (Danukusumo, 2017). Teknik pembelajaran mendalam memberikan rancangan arsitektur yang sangat kuat pada teknik *supervised learning*. Salah satu metode *deep learning* yang saat ini sedang berkembang adalah *convolutional neural network*. Jaringan tersebut terkenal dengan masukan berupa *image*, kemudian melalui *convolution layer* dan diolah dari filter yang sudah ditentukan, pada tiap *layer* tersebut mendapatkan hasil berupa pola dari beberapa bagian citra yang memudahkan proses *classification*. Pembelajaran mendalam merupakan model pendekatan berbasis *machine learning* dalam berbagai tugas diantaranya klasifikasi teks, termasuk analisis sentimen, kategorisasi berita, penjawab pertanyaan, dan inferensi bahasa alami.

Klasifikasi teks juga dikenal sebagai kategorisasi teks yang merupakan masalah umum dalam *natural language processing* (NLP), yang bertujuan untuk menetapkan label atau tag pada unit tekstual seperti kalimat, paragraf, dan dokumen. Data teks dapat berasal dari sumber yang berbeda, termasuk data web, email, pesan, media sosial, tiket, klaim asuransi, ulasan pengguna, dan tanya jawab dari pelanggan (Minaee et al., 2021).

Menurut (Bustami, 2014) dalam penelitiannya yang judulnya Penerapan Metode *Naive Bayes* Untuk Mengklasifikasi Data. Klasifikasi adalah proses untuk menemukan suatu model yang menjelaskan atau

membedakan suatu kelas data yang bertujuan untuk melakukan perkiraan pada kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam label yang berbeda didasarkan pada aturan atau fungsi tertentu. Secara umum metode *classification* dibagi menjadi dua macam yaitu klasifikasi *unsupervised* dan *supervised*. Metode klasifikasi *supervised* merupakan metode learning dengan bantuan. Sedangkan metode klasifikasi *unsupervised* merupakan kebalikan dari metode *supervised* (Hartono et al., 2020).

Cara kerja klasifikasi data terdiri dari dua proses yaitu proses pembelajaran di mana membuat model klasifikasi didasarkan kumpulan data yang sudah tersedia dan proses klasifikasi yaitu menentukan model yang dipakai untuk melakukan prediksi kelas untuk data yang diberikan. Pada proses pembelajaran data latih dianalisis dengan klasifikasi algoritma (Han et al., 2012). Pada Tabel 2.1 ini menjelaskan formula beserta tabel *confusion matriks* yang digunakan untuk pengukuran evaluasi kinerja klasifikasi.

Tabel 2.1 Formula Evaluasi Kinerja Klasifikasi

<i>Measure</i>	<i>Formula</i>
accuracy, recognition rate	$\frac{TP+TN}{P+N}$
error rate, misclassification rate	$\frac{FP+FN}{P+N}$
sensitivity, true positive rate, recall	$\frac{TP}{P}$
specificity, true negative rate	$\frac{TN}{N}$
precision	$\frac{TP}{TP+FP}$
$F$ , $F_1$ , $F$ -score, harmonic mean of precision and recall	$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$
$F_\beta$ , where $\beta$ is a non-negative real number	$\frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$

Terdapat empat istilah yang terdapat pada Tabel 2.1 yang perlu kita ketahui yang digunakan dalam menghitung ukuran evaluasi kinerja klasifikasi. Dibawah ini akan diberikan penjelasan untuk memahami istilah pada Tabel 2.1.

- *True positives* (TP): Data yang mendapat hasil klasifikasi dengan benar sesuai dengan data positif.
- *True Negatives* (TN): Data yang mendapat hasil klasifikasi dengan benar sesuai dengan data yang negatif.
- *False Positives* (FP): Data yang mendapat hasil klasifikasi kurang tepat dari data positif.
- *False Negatives* (FN): Data yang mendapat hasil klasifikasi dengan kurang tepat.

Dari formula evaluasi kinerja klasifikasi yang sudah dijelaskan diatas didapatkan nilai dari *confusion matrik* pada Tabel 2.2.

Tabel 2.2 *Confusion Matriks*

		Predicted class		Total
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Sumber : (Han et al., 2012)

*Confusion matrix* adalah suatu metode yang menganalisis tingkat seberapa baik model *classification* yang dilakukan dapat mengenali data baru yang berasal dari kelas yang berbeda.

### 2.2.2 *Natural Language Processing (NLP)*

*Natural Language Processing (NLP)* adalah metode komputerisasi untuk melakukan analisis text yang berdasarkan pada seperangkat teori dan seperangkat teknologi. Dan, sebagai bidang penelitian dan pengembangan yang sangat aktif (Liddy, 2001). Selain itu, NLP merupakan suatu bidang perancangan metode dan algoritma yang menerima input atau menghasilkan output data bahasa alami yang tidak terstruktur. Komposisi Bahasa itu sendiri yaitu : huruf membentuk kata, kata membentuk suatu frasa dan kalimat. Makna suatu frasa bisa lebih besar daripada makna kata-kata individual yang menyusunnya, dan mengikuti seperangkat aturan yang rumit. Untuk menafsirkan sebuah teks, kita perlu bekerja di luar tingkat huruf dan kata, dan melihat rangkaian kata yang Panjang seperti kalimat, atau bahkan dokumen lengkap (Hirst et al., 2015).

Tujuan *Natural language processing (NLP)* digunakan menggambarkan fungsi dari komponen suatu *software* atau *hardware* pada sistem yang menganalisa atau mensintesis bahasa lisan ataupun tulisan. *Natural language processing (NLP)* berfokus pada pengelolaan bahasa natural yaitu bahasa yang sering digunakan oleh manusia untuk berkomunikasi satu sama lain, seperti matematika atau notasi, atau bahasa komputer, seperti Java, LISP, dan C++. Istilah *natural language processing (NLP)* pada awalnya disebut sebagai *natural language understanding (NLU)* yang bertujuan untuk memiliki sistem komputer yang benar-benar

memahami bahasa alami semirip mungkin dengan manusia (Peter Jackson & Isabelle Moulinier, 1984). Sementara itu NLP dapat dikatakan *true* NLU, jika sistem NLU dapat melakukan beberapa hal berikut ini :

1. Parafrase masukan text
2. Terjemah text
3. Menjawab pertanyaan tentang isi text
4. Mengambil kesimpulan dari text

Disamping itu, *Natural language processing* (NLP) adalah bidang penelitian dan aplikasi yang relatif baru, dibandingkan dengan yang lain. Pendekatan teknologi informasi NLP menyediakan teori dan implementasi untuk berbagai aplikasi. Implementasi NLP yang paling sering digunakan adalah sebagai berikut : *Information Retrieval, Information Extraction (IE), Question-Answering, Summarization, Machine Translation, and Dialogue System* (Liddy, 2001).

### **2.2.3 Preprocessing**

*Preprocessing* digunakan untuk membuat kinerja algoritma klasifikasi agar maksimal. Pada umumnya terdapat empat tahapan *preprocessing* yang digunakan yaitu *case folding, stopwords removal, tokenizing* dan *stemming*. *Case folding* dengan cara membuat teks pada dataset menjadi bentuk *lowercase* atau bisa disebut proses mengubah semua dataset menjadi suatu bentuk standar (biasanya huruf kecil) (Mahendra et al., 2020). *Tokenizing* dilakukan dengan cara memisah kalimat menjadi satuan kata. *Stopword* merupakan perincihan dari analisis hasil dari kata dalam Bahasa Indonesia,

yang berisi kata-kata Bahasa Indonesia yang paling sering yang sering muncul namun tidak memiliki arti tertentu seperti kata ‘merupakan’, ‘yaitu’, ‘adalah’, ‘ada’, ‘maupun’, dan sebagainya. Apabila kata-kata pada kamus ditemukan pada dataset, maka kata tersebut akan dihapus. *Stemming* berfungsi untuk memfilter *root word* dari setiap kata yang terdapat pada dataset. Sehingga jika terdapat kata yang berbeda akan tetapi serupa misal seperti ‘pengukur’ dan ‘pengukuran’ dianggap satu kata yang sama yaitu ‘ukur’. *Stemming* dokumen Bahasa Indonesia sesuai PUEBI paling populer dikembangkan dengan menggunakan algoritma Nazief karena menghasilkan presisi yang lebih tinggi dibanding algoritma yang lain. Dari algoritma Nazief, dikembangkan sebuah *library stemming* dalam bentuk Bahasa Indonesia yang bernama Sastrawi (Ma’rifah et al., 2020). *Preprocessing* biasanya digunakan untuk meningkatkan efisiensi dan akurasi model dengan mengurangi variasi yang tidak dimodelkan. Pilihan pra-pemrosesan yang sesuai atau kombinasi metode *preprocessing* dapat mempengaruhi kinerja analisis dan penyalahgunaan teknik *preprocessing* dapat menurunkan kinerja model. Teknik-teknik pada *preprocessing* ini tidak saling eksklusif; mereka mungkin bekerja sama. Misalnya, pembersihan data dapat melibatkan transformasi untuk memperbaiki data yang salah, seperti dengan mengubah semua entri untuk bidang tanggal ke format umum (Han et al., 2012)

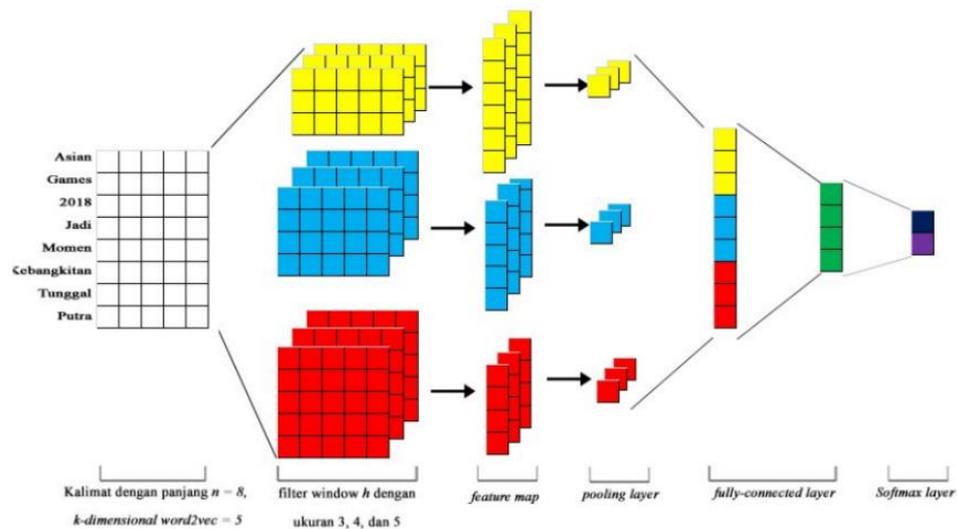
#### **2.2.4 Convolutional Neural Network (CNN)**

*Convolutional Neural Network* atau yang disingkat CNN adalah *neural network multilayer* yang terdiri lebih dari dua *deep layer* dan dihubungkan dengan lapisan *fully connected* mirip jaringan saraf *multilayer*. Jaringan ini sangat baik pada masalah *machine learning*. Khusus pada sistem yang mempunyai hubungan dengan data *image*, seperti kumpulan data *image classification*, *computer vision*, dan dalam NLP dan hasil yang didapatkan sangat bagus (Saad ALBAWI et al., 2017).

Secara umum metode *convolutional neural network* tidak jauh beda dari jaringan saraf biasanya, yaitu terdiri oleh *neuron* yang mempunyai *bias*, *weight* dan fungsi aktivasi. Tugas dari algoritma CNN merupakan sebuah arsitektur yang bisa mengenali suatu prediksi data dari suatu objek seperti text, potongan suara, dan *image*.

*Convolutional neural network* adalah penerapan pembelajaran mendalam yang terdapat perbedaan dengan metode pembelajaran mesin. Dapat diperhatikan pada waktu yang dilakukan untuk eksekusi yang akan dijalankan, model pada *deep learning* menggunakan estimasi waktu yang relatif lama dari pada *machine learning*. Selain itu, dari segi perangkat keras harus menggunakan spesifikasi yang lebih tinggi untuk perhitungan operasi matriks dalam jumlah besar. *Machine learning* cocok untuk data yang jumlahnya tidak terlalu banyak atau data kecil, sedangkan model *deep learning* bagus untuk digunakan pada data yang lebih besar. Kontribusi paling utama pada algoritma ini adalah pada layer konvolusi dan lapisan

*pooling*. Pada Gambar 2.1 akan dijelaskan mengenai arsitektur model *deep convolutional neural network*.



Gambar 2.1 Model *Deep Convolutional Neural Network*

*Convolutional Neural Network* sebagai tugas *natural language processing*, masukannya berupa kalimat yang direpresentasikan dalam bentuk *matriks*. Penjelasan yang terdapat pada Gambar 2.1, Jika  $x_i \in \mathbb{R}^k$  merepresentasikan *k-dimensional word embedding* sesuai kata ke  $i$  pada *sentence*, kalimat dengan panjang  $n$  dan dituliskan menjadi fungsi pada persamaan (1).

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_n \quad (1)$$

Operator  $\oplus$  adalah operator yang digunakan untuk menggabungkan text yang telah dirubah menjadi *vector* kedalam bentuk matrik. Operasi pada konvolusi menggunakan filter  $w \in \mathbb{R}^{h \times k}$ , dengan *size window*  $h$  untuk jumlah urutan kata yang diproses. Jendela yang digunakan penelitian diatas

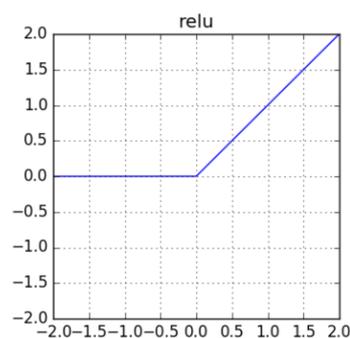
adalah 3, 4, dan 5. Fitur  $C_i$  didapatkan dari *window* ukuran  $h$  dan melakukan *dot product* dengan matrik masukan berupa  $x_{i:i+h-1}$  dan dilanjutkan *activation function*. Berikut merupakan beberapa fungsi aktivasi :

- *Sigmoid*

*Sigmoid* yaitu menerima angka tunggal untuk mengubah nilai  $x$  menjadi sebuah nilai yang memiliki range mulai dari 0 hingga 1 (Sitepu & Sigiro, 2021).

- ReLU (*Rectifier Linear Unit*)

ReLU merupakan *activation function* yang memiliki perhitungan paling sederhana. Apabila elemen bernilai *negative* maka nilainya diset menjadi 0, pada fungsi aktivasi ini tidak ada operasi perkalian atau pembagian yang digunakan. Kelebihannya adalah waktu yang dilakukan pada pelatihan dan pengujian lebih cepat.

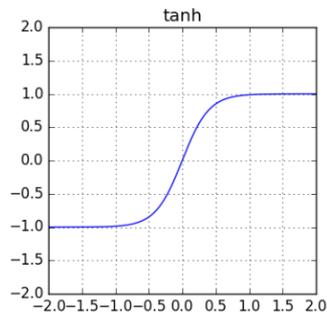


Gambar 2.2 Struktur pada ReLU

- Tanh (*Hyperbolic tangent function*)

Tanh merupakan fungsi aktivasi yang biasanya lebih cepat mencapai konvergensi dibandingkan fungsi aktivasi *sigmoid* dan akan

menghasilkan akurasi yang lebih tinggi. Pada performa fungsi aktivasi Tanh mirip dengan performa *classification* yang dihasilkan oleh *activation function ReLU* (Wibawa, 2017).



Gambar 2.3 Struktur pada Tanh

Pada fungsi *rectifier linear unit* atau ReLU direpresentasikan oleh persamaan (2).

$$c_i = \max(0, w \cdot x_{i:h-1}) \quad (2)$$

Setiap filter melakukan konvolusi pada *sentence matriks*  $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$  dan menghasilkan *feature-maps* dengan  $c \in \mathbb{R}^{n-h+1}$ . Fungsi *feature map* direpresentasikan oleh persamaan (3).

$$c = [c_1, c_2, \dots, c_{n-h+1}] \quad (3)$$

Untuk meminimalisir model yang terjadi *overfitting* harus menggunakan lapisan *dropout* yang berfungsi membuang *neuron* secara acak pada proses pelatihan (Srivastava et al., 2014). Probabilitas lapisan *dropout* yang dipakai pada penelitian diatas sebesar 0,5. Pada tahap *training* juga melakukan *pooling* pada *feature map*. Metode *pooling* yang paling umum digunakan adalah *MaxPooling*, yaitu mengambil nilai tertinggi atau  $\hat{c} = \max\{c\}$  sebagai fitur berdasarkan suatu filter. Tujuan dari *pooling* adalah

untuk mendapatkan fitur yang penting untuk mewakili fitur lain pada setiap *feature map*. Fitur dihasilkan dari *pooling* digunakan untuk proses klasifikasi pada *fullyconnected layer*. Pada tiap-tiap fitur akan dihubungkan dengan setiap neuron pada lapisan *fullyconnected* menggunakan bobot independen dan jumlah neuron sebanyak 50. Pada tiap neuron dari *fully connected layer* akan menjadi input bagi lapisan *softmax* (Juwiantho et al., 2020).

Untuk menemukan model *Convolutional Neural Network* yang ideal dan optimal harus melakukan *trial* dan *error* dengan mengubah sejumlah parameter pada model salah satunya dengan metode optimasi. Salah satu metode optimasi adalah *adam*. Algoritma *adam* adalah algoritma optimisasi stokastik berdasarkan pendekatan adaptif momen order rendah yang dapat dengan mudah diimplementasikan karena mempunyai perhitungan yang efisien. Selain itu juga mempunyai jejak memori yang kecil, tidak berubah pada pemberian skala gradien dan cocok diterapkan pada data atau parameter dengan jumlah yang besar (Wibawa, 2017).

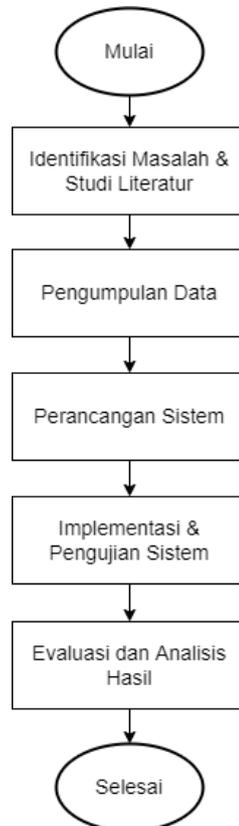
## BAB III

### METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang metodologi penelitian yang akan digunakan pada penelitian ini. Metodologi penelitian ini dirancang sebagai panduan penulis agar penelitian tugas akhir ini berjalan secara sistematis dan terstruktur.

#### 3.1 Tahapan Penelitian

Dalam melakukan penelitian ini, terdapat tahapan penelitian untuk menyelesaikan masalah terkait klasifikasi kalimat perbincangan masyarakat pada masa pandemi covid-19 menggunakan media sosial dengan metode *Convolutional Neural Network* (CNN). Pada tahapan dari penelitian ini direpresentasikan pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Berdasarkan tahapan penelitian pada Gambar 3.1 dapat dilihat langkah-langkah yang dilakukan untuk mendapatkan hasil dari penelitian ini. Tahapan pertama yang dilakukan pada penelitian tugas akhir ini adalah melakukan identifikasi masalah terkait objek dan metode penelitian yang digunakan, identifikasi masalah merupakan upaya untuk mengenali apa masalah yang ada dan bagaimana mendefinisikan suatu permasalahan yang berasal dari rumusan masalah yang sudah dibuat sebelumnya. Setelah Identifikasi masalah selesai selanjutnya adalah melakukan kegiatan studi literatur dengan menggunakan referensi dari buku terkait topik penelitian dan juga dari penelitian yang sudah dilakukan sebelumnya. Tujuan dari identifikasi masalah adalah mencari informasi yang relevan dengan objek dan metode yang diteliti dan juga memperdalam pengetahuan tentang topik yang diteliti.

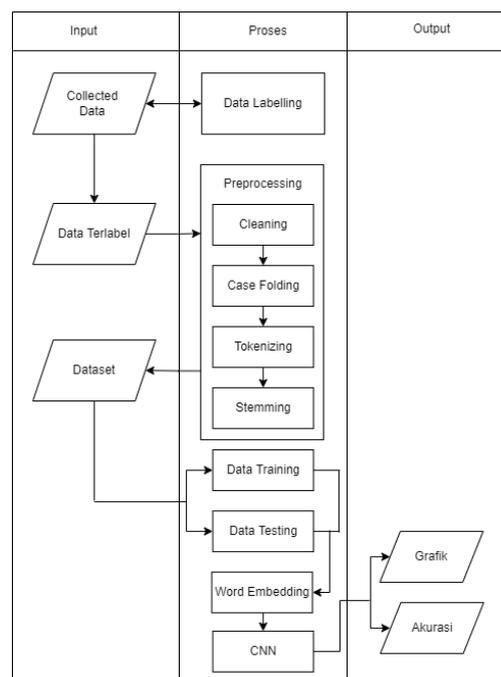
Tahapan selanjutnya yang harus dilakukan adalah pengumpulan data. Tahapan pengumpulan data atau *collecting data* pada penelitian ini bertujuan untuk mendapatkan data yang digunakan terkait penelitian ini, data yang digunakan pada penelitian ini berupa data primer. Data primer itu sendiri merupakan data utama yang akan diolah dimana outputnya untuk mendapatkan hasil penelitian yang dilakukan. Teknik pengumpulan data yang dapat dipakai untuk pengumpulan informasi kuantitatif, yaitu kuesioner, wawancara terencana, dan observasi.

Setelah dilakukan pengumpulan data, tahap selanjutnya adalah perancangan sistem yang bertujuan untuk mendefinisikan sistem yang akan dibuat pada penelitian ini. Setelah perancangan sistem dibuat, berikutnya adalah mengimplementasikan dan melakukan pengujian algoritma dan metode yang

digunakan pada sistem ini, dimana pada penelitian ini menggunakan salah satu algoritma *deep learning* yaitu *convolutional neural network* (CNN). Pada tahapan akhir penelitian dilakukan analisis dan evaluasi dari hasil performa pengujian algoritma dari proses klasifikasi kalimat perbincangan masyarakat pada masa pandemi covid-19 dengan algoritma *Convolutional Neural Network* (CNN).

### 3.2 Perancangan Sistem

Sebelum membangun sistem yang dibuat untuk penelitian ini, harus ada sebuah desain dari perancangan sistem yang akan dibangun. Dari desain perancangan sistem itu akan dijabarkan proses program dibangun mulai dari tahap pengumpulan data, lalu dilanjut dengan *preprocessing* yang memiliki 4 tahapan yaitu *Cleaning*, *Case Folding*, *Tokenizing* dan *Stemming*. Setelah *preprocessing* dilakukan selanjutnya ada proses training model CNN, dimana akan didapatkan suatu model CNN yang digunakan untuk klasifikasi teks. Untuk menjelaskan proses perancangan sistem pada penelitian ini, maka dijelaskan pada Gambar 3.2.



Gambar 3.2 Desain Sistem

### 3.3 *Collecting Data*

Pada tahapan ini dilakukan proses mengumpulkan data atau *collecting data*. Proses pengumpulan data ini mempunyai tujuan untuk mendapatkan data yang valid guna untuk mencapai tujuan penelitian. Data yang dipakai pada penelitian ini berupa kalimat yang di ambil dari *tweet* pada media sosial *Twitter*. Proses pengumpulan data yaitu dengan cara *crawling* dengan menggunakan API yang disediakan oleh platform media sosial *Twitter* menggunakan Bahasa pemrograman python. Pada proses *crawling* data ini peneliti mengambil *tweet* berbahasa Indonesia dengan menggunakan kata kunci “Pandemi Covid-19” dengan pembatasan waktu mulai bulan September 2020 hingga bulan September 2021. Setelah data terkumpul maka data tersebut disimpan dalam bentuk excel yang akan diolah dalam tahap selanjutnya. Sebelum melakukan proses selanjutnya, data yang akan dijadikan *dataset* tersebut dilakukan *remove duplicate* terlebih dahulu karena untuk membersihkan data *tweet* yang kemungkinan sama dan data tersebut yang akan digunakan dalam proses pengklasifikasian.

Setelah data *tweet* dari platform media sosial *Twitter* didapatkan maka langkah selanjutnya yaitu pelabelan oleh pakar secara manual. Dalam tahap ini yang dimaksud pelabelan oleh pakar adalah mengelompokkan jenis *tweet* pada data dimana pakar tersebut merupakan seseorang yang ahli dan kompeten dalam bidang bahasa karena yang dikelompokkan merupakan perbincangan masyarakat dan seorang pakar tersebut mengelompokkan data yang dibagi menjadi kelas positif dan negatif. Berikut merupakan beberapa data kalimat perbincangan masyarakat yang sudah dilakukan tahap pelabelan manual oleh pakar dalam bidang bahasa.

Tabel 3.1 Data Terlabel

<i>Tweet</i>	<i>Label</i>
RT @ernestprakasa: Pak Luhut Yth. yang juga sangat menyiksa batin dari penanganan COVID-19 adalah vonis ringan bagi koruptor bansos. 😊Luh...	Negatif
RT @airasia_indo: Pelanggan AirAsia yang terhormat, sebagai upaya membantu pemerintah menangani pandemi Covid-19 & sejalan dengan perpanjian...	Positif
"Kadang ada juga orang yang ditangkap atau kriminalisasi dan juga dianggap Positif Covid-19. Padahal tidak terbukti... <a href="https://t.co/OGRD3SmFL5">https://t.co/OGRD3SmFL5</a>	Negatif
Dalam rangka penanganan pandemi Covid-19 Kemendikbudristek pada tahun 2020 dan 2021 telah menyalurkan bantuan... <a href="https://t.co/6wFL6ZGyuQ">https://t.co/6wFL6ZGyuQ</a>	Positif
HMI mau kepung istana mulai 6 Agt, puncaknya 16 Agt 2021 dgn alasan pemerintah gagal atasi covid-19. Org2 gagal dlm... <a href="https://t.co/o7Km339gqU">https://t.co/o7Km339gqU</a>	Negatif
Jumlah pengguna dompet digital (e-wallet) pada masa pandemi Covid-19 semakin meningkat di 2021. Berdasarkan data... <a href="https://t.co/ClyNZx7qRF">https://t.co/ClyNZx7qRF</a>	Positif
Percenatan vaksinasi Covid-19 untuk anak akan memberikan peluang lebih cepat sekolah dibuka kembali di masa pandemi. <a href="https://t.co/8sgIISYV6c">https://t.co/8sgIISYV6c</a>	Positif
"Kadang ada juga orang yang ditangkap atau kriminalisasi dan juga dianggap Positif Covid-19. Padahal tidak terbukti... <a href="https://t.co/OGRD3SmFL5">https://t.co/OGRD3SmFL5</a>	Negatif
RT @hariankompas: Setelah muncul kasus pungutan liar bantuan sosial, terkuak pula ada anggaran bahan pakaian wakil rakyat di DPRD Kota Tang...	Negatif
Panglima TNI mengingatkan tenaga kesehatan yang tengah berjuang menanggulangi pandemi Covid-19 untuk selalu meneran... <a href="https://t.co/sU8IdZelrU">https://t.co/sU8IdZelrU</a>	Positif

### 3.4 Preprocessing

Tahap yang dilakukan selanjutnya adalah *text preprocessing* atau pra-pemrosesan teks. Output pada tahap *preprocessing* ini berupa *dataset* yang

selanjutnya akan dibagi menjadi data latih dan data uji. Data latih merupakan data pembelajaran yang digunakan untuk melatih algoritma klasifikasi yang digunakan. Sedangkan, data uji merupakan data yang di uji coba pada machine learning. Berikut tahapan-tahapan pada pra-pemrosesan teks.

### 3.4.1 *Cleaning*

Pada penelitian ini tahapan *preprocessing* yang dilakukan pertama adalah proses *cleaning*. Pada tahapan ini menghapus atau membersihkan tautan, karakter, emoticon, dan tanda baca yang terdapat pada teks *tweet*. Tanda baca yang harus dihilangkan yaitu seperti (!'#\$%&'[]\*+,-./:;<=>?@[\\]^\_`{|}~), menghilangkan simbol karakter atau *emoticon* seperti ( 😊 😏 😂 🗿 ), Menghapus tautan atau URL seperti (“https://t.co/eOfeu0nSqR”), menghapus *mention* pengguna twitter yang menggunakan awalan (@), menghapus *hashtag* (#), menghapus *retweet* (rt), dan yang terakhir menghapus spasi dan enter berlebih. Dalam tahapan ini ditunjukkan contoh pada Tabel 3.3.

Tabel 3.2 *Cleaning*

Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>
RT @ernestprakasa: Pak Luhut Yth, yang juga sangat menyiksa batin dari penanganan COVID-19 adalah vonis ringan bagi koruptor bansos. 😏 Luh...	Pak Luhut Yth yang juga sangat menyiksa batin dari penanganan COVID19 adalah vonis ringan bagi koruptor bansos Luh

### 3.4.2 *Case Folding*

*Case Folding* merupakan suatu cara pada *preprocessing* yang berfungsi untuk mengganti huruf kapital kedalam bentuk huruf kecil. Pada *dataset*

yang digunakan pada penelitian ini yang semula terdapat huruf kapital akan diubah sehingga seluruh teks menjadi huruf kecil. Seperti yang terdapat pada Tabel 3.3.

Tabel 3.3 *Case Folding*

<b>Sebelum <i>Case Folding</i></b>	<b>Setelah <i>Case Folding</i></b>
Pak Luhut Yth yang juga sangat menyiksa batin dari penanganan COVID19 adalah vonis ringan bagi koruptor bansos Luh	pak luhut yth yang juga sangat menyiksa batin dari penanganan covid19 adalah vonis ringan bagi koruptor bansos luh

### 3.4.3 *Tokenizing*

*Tokenizing* merupakan suatu tahapan pada *preprocessing* yang berfungsi untuk memecah kalimat menjadi setiap kata, proses ini digunakan agar mendapatkan sebuah kata yang pada tahap selanjutnya diubah menjadi vektor. Seperti yang terdapat pada Tabel 3.4.

Tabel 3.4 *Tokenizing*

<b>Sebelum <i>Tokenizing</i></b>	<b>Setelah <i>Tokenizing</i></b>
pak luhut yth yang juga sangat menyiksa batin dari penanganan covid19 adalah vonis ringan bagi koruptor bansos luh	“pak”, “luhut”, “yth”, “yang”, “juga”, “sangat”, “menyiksa”, “batin”, “dari”, “penanganan”, “covid19”, “adalah”, “vonis”, “ringan”, “bagi”, “koruptor”, “bansos”, “luh”

### 3.4.4 *Stemming*

*Stemming* adalah proses mengubah kata yang terdapat pada *dataset* ke dalam bentuk kata dasar (*root word*) dengan cara menggunakan aturan tertentu. Pada teks Bahasa Indonesia proses untuk menghilangkan imbuhan yang ada pada suatu kata (*affixes*). Baik kata imbuhan yang ada diawal kata

(*prefixes*), imbuhan yang ada ditengah atau sisipan (*infixes*), imbuhan yang ada diakhir kalimat (*suffixes*), ataupun imbuhan kombinasi yang ada dari awalan dan imbuhan akhiran (*confixes*). Contoh proses *Stemming* dijabarkan pada table 3.5

Table 3.5 *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
pak”, “luhut”, “yth”, “yang”, “juga”, “sangat”, “menyiksa”, “batin”, “dari”, “penanganan”, “covid19”, “adalah”, “vonis”, “ringan”, “bagi”, “koruptor”, “bansos”, “luh”	“pak”, “luhut”, “yth”, “yang”, “juga”, “sangat”, “siksa”, “batin”, “dari”, “tangan”, “covid19”, “adalah”, “vonis”, “ringan”, “bagi”, “koruptor”, “bansos”, “luh”

### 3.5 *Word Embedding*

Pada tahapan ini dilakukan *word embedding*. *Word embedding* merupakan metode yang digunakan untuk membuat sebuah vektor diubah dari representasi kata yang memungkinkan kata-kata dengan arti yang sama untuk memiliki representasi yang serupa. (Brownlee, 2017). Salah satu teknik pada *word embedding* yang dapat digunakan adalah *embedding layer*. *Embedding layer* merupakan teknik *word embedding* yang dapat dipelajari dengan model *neural network* pada permasalahan *natural language processing*. Seperti permodelan bahasa dan klasifikasi dokumen.

Table 3.6 Kalimat Negatif

<b>Tweet Negatif</b>	“siksa”, “batin”, “dari”, “tangan”, “covid19”
----------------------	---

Cara kerja *embedding layer* mengharuskan sebuah data yang telah melewati tahapan *preprocessing* terlebih dahulu, selanjutnya kalimat dipecah menjadi satuan

kata seperti pada tabel 3.7. Selanjutnya kata tersebut diberikan nilai vektor atau bobot yang diinisialisasi dengan angka acak kecil (Brownlee, 2017). Berikut merupakan simulasi dari teknik *embedding layer* pada Tabel 3.7.

Table 3.7 Vektor Kata

Kalimat	Vektor				
siksa	0,5	0,2	-0,3	-0,1	-0,2
batin	-0,1	-0,3	0,2	0,4	0,1
dari	0,3	-0,3	0,1	0,1	0,2
tangan	0,2	-0,3	0,4	0,2	-0,1
covid19	0,1	0,2	-0,1	-0,1	-0,3

### 3.6 Proses Perhitungan Metode *Convolutional Neural Network*

Pada tahapan perhitungan dengan metode *Convolutional Neural Network* ini, data yang sudah mendapatkan bobot dari proses *word embedding* selanjutnya diolah menggunakan metode tersebut. Pada proses perhitungan metode CNN terdapat 2 lapisan yang dilakukan yaitu *feature learning layer* dan *Classification layer*. Berikut ini adalah simulasi dari alur perhitungan metode *Convolutional Neural Network*.

#### 3.6.1 *Feature Learning Layer*

Pada *feature learning layer* terdapat konvolusi. Konvolusi yaitu menggabungkan dua buah baris angka yang menghasilkan baris angka ketiga. Pada penelitian ini, dua buah baris angka terdapat pada *input* dan *kernel*, kombinasi dari dua baris angka tersebut membuat baris ketiga yang disebut *output*. *Input*, *kernel* dan *output* tersebut memiliki deret bilangan

yang berbentuk matriks. *Input* disini merupakan hasil dari tahap *word embedding*, dan *kernel* digunakan sebagai filter untuk data *input* sehingga menghasilkan *feature-map* (Mahendra et al., 2020). Untuk dapat lebih memahami proses konvolusi, peneliti membuat simulasi dengan deret angka pada *input* berukuran 5x5 dan menggunakan *filter* berukuran 3x3 dan menghasilkan *output* berukuran 3x3, seperti yang terdapat pada Gambar 3.3.

0,5	0,2	-0,3	-0,1	-0,2
-0,1	-0,3	0,2	0,4	0,1
0,3	-0,3	0,1	0,1	0,2
0,2	-0,3	0,4	0,2	-0,1
0,1	0,2	-0,1	-0,1	-0,3

 $\times$ 

-1	0	1
-1	0	1
-1	0	1

 $=$ 

-0,7	0,8	0,1
0,3	1,6	-0,5
-0,2	0,6	-0,6

Gambar 3.3 Proses Perhitungan Konvolusi

Penjelasan dari Gambar 3.3 adalah perhitungan dengan cara penggeseran matrik pada *input* dikalikan dengan *filter* yang menghasilkan sebuah *output* yang dihitung sebagai berikut :  $(0,5 \times (-1)) + (0,2 \times 0) + ((-0,3) \times 1) + ((-1) \times (-1)) + ((-0,3) \times 0) + (0,2 \times 1) + (0,3 \times (-1)) + ((-0,3) \times 0) + (0,1 \times 1) = -0,7$ .

Tahap Perhitungan selanjutnya adalah *activation function* berfungsi untuk mendapatkan nilai *non-linier* pada hasil konvolusi. Pada perhitungan *Activation Function* ini diberi fungsi *ReLU* untuk menghindari nilai negatif, dengan  $x$  adalah nilai yang didapatkan dari *featuremaps*. Rumus *ReLU* yang digunakan ialah  $f(x) =$

$\max(x, 0)$  dengan  $x$  merupakan *input* (Mahendra et al., 2020). Untuk dapat memahami ReLU dapat disimulasikan pada Gambar 3.4.

-0,7	0,8	0,1	=	0	0,8	0,1
0,3	1,6	-0,5		0,3	1,6	0
-0,2	0,6	-0,6		0	0,6	0

Gambar 3.4 Proses *Activation Function (ReLU)*

Setelan proses *activation function* dilakukan, selanjutnya terdapat tahap *pooling layer*. Pada *pooling layer*, data dikumpulkan dengan cara menggeser jendela dan melintasi *feature-maps*, kemudian menerapkan *non-linear* pada data yang terdapat di dalam jendela. Tujuan *Pooling layer* adalah untuk mereduksi dimensi dari *feature-maps* yang digunakan pada layer selanjutnya. *Pooling layer* ini menetapkan bahwa jaringan hanya berfokus terhadap pola yang paling penting (Nadhifa Ayu Shafirra & Irhamah, 2020). Untuk lebih memahami *pooling layer* dapat diilustrasikan dengan gambar 3.5.

0	0,8	0,1	=	0,8
0,3	1,6	0		1,6
0	0,6	0		0,6

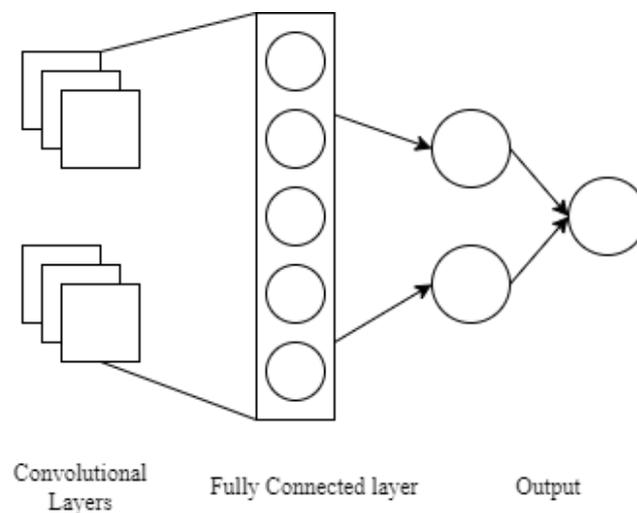
Gambar 3.5 Proses *Pooling Layer*

Perhitungan *pooling layer* pada Gambar 3.5 menggunakan metode *max-pooling* yaitu dengan cara mengambil nilai maksimum pada suatu jendela tertentu.

### 3.6.2 Classification Layer

Pada *classification layer* atau lapisan klasifikasi menggunakan *fully connected layer* yang terdiri dari tiga layer yaitu *input layer*, *hidden layer* dan *output layer*, jika diperhatikan pada layer ini sama seperti di model arsitektur ANN. Oleh karena itu beberapa penelitian menyebutkan bahwa *fully connected layer* ini sebagai ANN.

*Fully connected layer* atau lapisan neuron yang terhubung sepenuhnya mengambil vektor sebelumnya sebagai salah satu parameter dari jumlah kelas yang akan diprediksi (negatif dan positif). Pada *hidden layer* juga terdapat *activation function*, fungsi aktivasi yang biasa digunakan yaitu ReLU, demikian juga pada *output layer* terdapat fungsi aktivasi, dalam kasus klasifikasi (yang lebih dari 2 label/kelas/kategori) yang umum digunakan adalah *Softmax*. Tujuan utama dari lapisan yang terhubung penuh atau *fully connected layer* merupakan untuk memproses data sehingga dapat diklasifikasikan (Raihan et al., 2020). Pada gambar 3.6 di bawah ini adalah gambaran bagaimana lapisan yang terhubung penuh dibangun.



Gambar 3.6 Contoh *fully connected layer*

Pada gambar 3.6 merupakan contoh *fully connected layer* yang terhubung ke semua neuron dan bersama-sama memprediksi output dari dua kelas. Fungsi *softmax* yang terdapat di output layer pada tahapan *fully connected/classification* tujuannya untuk menghitung probabilitas suatu kelas (Kristoffer & Hagelbäck, 2017). Pada tahapan ini juga menggunakan *dropout*. *Dropout* adalah teknik untuk mereduksi *overfitting* dimana terdapat beberapa *neuron* dipilih secara acak dan tidak digunakan selama pelatihan (dengan kata lain: menghapus fitur/mengurangi variansi). Selain mengurangi terjadinya *overfitting*, penambahan *dropout* juga mempercepat tahapan pelatihan pada model (Raihan et al., 2020).

### 3.7 Skenario Pengujian

Tahapan ini merupakan tahapan akhir sistem yaitu dengan melakukan pengujian terhadap kinerja suatu sistem yaitu dengan cara menghitung *accuracy*, *Recall*, Presisi, dan *F1-Score*. Menghitung performa *classification* pada penelitian ini yaitu dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan metode yang dipakai untuk mengevaluasi dan menyajikan hasil klasifikasi tentang data aktual dan data prediksi dari klasifikasi yang telah dilakukan. Konsep pengukuran kinerja *confusion matrix* dapat dilihat pada Tabel 3.8.

Tabel 3.8 *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	True Positive (TP)	False Positive (FP)
Negatif	False Negative (FN)	True Negative (TN)

Berikut penjelasan pada Tabel 3.8 dimana *True Positif* (TP) adalah data yang mendapat hasil klasifikasi dengan benar sesuai dengan data positif. TN atau *True Negative* adalah data yang mendapat hasil klasifikasi dengan benar sesuai dengan data yang negatif. *False Positif* (FP) adalah data yang mendapat hasil klasifikasi kurang tepat dari data positif. Dan FN (*False Negative*) merupakan suatu data yang mendapat hasil klasifikasi kurang tepat. Nilai pada setiap *confusion matrix* digunakan untuk menghasilkan nilai *accuracy*, presisi, *recall* dan *F1-Measure* yang ditunjukkan pada rumus dibawah ini.

$$Akurasi (\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Precision (\%) = \frac{TP}{TP + FP} \times 100\%$$

$$Recall (\%) = \frac{TP}{TP + FN} \times 100\%$$

$$F1 - Score(\%) = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%$$

Akurasi adalah tingkat kedekatan antara nilai *predict* dengan nilai *actual* atau nilai sebenarnya, *Precision* merupakan tingkat kesesuaian pada informasi yang diinginkan dengan jawaban dari model klasifikasi, *Recall* adalah data penghapusan yang telah berhasil didapatkan dari data yang berkaitan dengan kata kunci atau tingkat keberhasilan pada sistem untuk menemukan kembali suatu informasi dan *F-Measure* atau yang biasa disebut *F1-Score* adalah perhitungan evaluasi dari hasil kombinasi antara *Recall* dan *Precision* (Mahendra et al., 2020).

## BAB IV

### IMPLEMENTASI DAN PEMBAHASAN

#### 4.1 Skenario Uji Coba

Skenario uji coba merupakan tahapan yang digunakan dalam pengujian pada sistem yang sudah dibangun pada penelitian ini. Pada skenario uji coba ini dilakukan untuk menghitung akurasi, presisi, recal, dan *F1-Score* dengan menggunakan Bahasa pemrograman *python*. Tahapan yang dilakukan sebelum menghitung akurasi, presisi, recal, dan *F1-Score* adalah melakukan tahapan proses *collecting data*, *preprocessing* dan proses klasifikasi dengan *Convolutional Neural Network* yang akan dijelaskan dibawah ini.

##### 4.1.1 Collecting Data

*Collecting data* atau pengumpulan data *tweet* pada media sosial twitter dengan cara *crawling* menggunakan *library tweepy*, dengan menggunakan *library* tersebut semua data dapat dikumpulkan kedalam satu file excel.

```
import tweepy
```

Gambar4.1 *Library Tweepy*

Pada pengumpulan data dari media sosial twitter membutuhkan akses *API* twitter yang dapat diambil dengan mendaftar menjadi *Twitter Developers* melalui website <https://developer.twitter.com/>

```
api_key = "sYnJuBiAC9IBXk4NL9FI98SXA"  
api_secret_key = "QBFVyKk0INsfYVRV7y1QxNsD0JnIG1cmGjtOzf  
9tJRVH7SBg0j"  
access_token = "1185398946539044864EpUf6Cd6Gpa764MKzcJ  
kX7DE5gLZC1"  
access_token_secret = "b0GVlqkBRb4sQGNCKZliuSTgtsw0XJ4d5  
wQZoOS6XCmyw"
```

Gambar4.2 *API twitter*

Dari tahapan *collecting data* dari media sosial dengan cara *crawling* didapatkan sebanyak 2.835 *tweet* yang selanjutnya dilakukan pemberian label pada setiap *tweet* berupa klasifikasi positif yang diberi kode 1 atau negatif yang diberi kode 0 oleh pakar dalam bidang bahasa. Setelah semua *tweet* dilabeli maka siap menjadi *dataset* untuk membangun model *Convolutional Neural Network*. Pada Tabel 4.1 merupakan jumlah data yang terlabel.

Tabel 4.1 Jumlah Data Terlabel

Klasifikasi	Kode	Jumlah
Positif	1	1.888
Negatif	0	947

#### 4.1.2 Preprocessing

Pada tahapan *preprocessing* ini, seluruh *tweet* pada *dataset* di proses dengan menggunakan beberapa proses yaitu *cleaning*, *case folding*, *tokenizing*, dan *stemming*. Berikut merupakan kode program dari *preprocessing*.

##### 3) Cleaning

Pada gambar 4.3 merupakan kode program yang digunakan untuk proses *cleaning* atau membersihkan data.

```
def cleanTxt(text):
    text = text.encode('ascii', 'ignore').decode()
    text = re.sub(r'https*\S+', ' ', text)
    text = re.sub(r'@\S+', ' ', text)
    text = re.sub(r'rt', ' ', text)
    text = re.sub(r#\S+', ' ', text)
    text = re.sub(r'\w+', ' ', text)
    text = re.sub(r'\w*\d+\w*', ' ', text)
    text = re.sub(r'\s{2,}', ' ', text)
    text = re.sub('rt', ' ', text)
```

```

text = re.sub(r"#", " <hash_tag> ", text)
text = re.sub(r'[\^\w\s]', '', text)
text = re.sub(r'<[^\<]+?>', '', text)
text = text.replace('\n', ' ')
text = re.sub(r'\s+', ' ', text)
return text
df['text'] = df['text'].apply(cleanText)

```

Gambar4.3 Kode Program *Cleaning*

- *Case Folding*

Pada gambar 4.4 merupakan kode program yang digunakan untuk proses *case folding* atau mengubah semua huruf kapital menjadi huruf kecil.

```
df['text'] = df['text'].str.lower()
```

Gambar4.4 Kode Program *Case Folding*

- *Tokenizing*

Pada gambar 4.5 merupakan kode program yang digunakan untuk proses *tokenizing* atau mengubah semua kalimat menjadi satuan kata.

```
df['text'] = df.apply(lambda row:nlTK.word_tokenize
(row['text']), axis=1)
```

Gambar4.5 Kode Program *Tokenizing*

- *Stemming*

Pada gambar 4.6 merupakan kode program yang digunakan untuk proses *stemming* atau mengubah semua suatu kata menjadi kata dasar. Pada proses *stemming* ini menggunakan *library* sastrawi untuk mereduksi kata dalam Bahasa Indonesia.

```

from Sastrawi.Stemmer.StemmerFactory
import StemmerFactory
import swifter
factory = StemmerFactory()

```

```

stemmer = factory.create_stemmer()
def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}
for document in df['text']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ''

print(len(term_dict))
print("-----")

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term, ":", term_dict[term])

print(term_dict)
print("-----")

def get_stemmed_term(document):
    return [term_dict[term] for term in document]
df['text2'] = df['text'].swifter.apply
    (get_stemmed_term)

```

Gambar 4.6 Kode Program *Stemming*

Selain menggunakan *library* Sastrawi, pada proses *stemming* juga menggunakan *library* Swifter yang digunakan untuk mempercepat proses *stemming* pada *dataframe*.

#### 4.1.2 Split Data

*Split data* pada tahapan ini yaitu membagi data menjadi data latih (*data train*) dan data uji (*data test*). Data latih atau *data train* merupakan data yang digunakan untuk melatih algoritma yang digunakan, sedangkan data uji atau *data test* merupakan data yang digunakan untuk menguji validasi dari algoritma yang digunakan. Perbandingan pembagian data latih dan data uji pada sistem ini yaitu 80% untuk data latih dan 20% untuk data uji dari 2.835 data yang digunakan.

Sebelum melakukan tahap *split data* ini, kata yang terdapat pada *dataset* dari tahap *preprocessing* harus diubah menjadi bentuk kata terindek dengan bantuan Tokenizer menggunakan Library Keras. Hasil dari perubahan menjadi bentuk numerik terdapat pada Tabel 4.2.

Tabel 4.2 Indek kata

Indeks kata	Kata
1	covid
2	pandemi
3	di
4	rt
5	dan
6	yang
7	untuk
8	tangan
9	ini
10	dalam
11	indonesia
12	bantu
13	agustus
14	tengah
15	jadi

Dari kata yang di indek terdapat 4.763 kata yang dihasilkan dari 2.835 data. Setelah kata berhasil di indek, Langkah selanjutnya adalah meng-encode

atau mengubah kalimat pada data menjadi nilai yang sesuai menggunakan fungsi `texts_to_sequences`. Pada gambar 4.7 merupakan hasil dari mengubah kalimat kedalam nilai yang sesuai.

```
[[852, 967, 486, 23, 7, 81, 487, 526, 68, 2, 1639, 443,
60, 968, 1120, 606, 3348], [1318, 757, 1319, 969, 109,
2118, 1320, 970, 971, 972, 1321, 758, 12, 3349], [1322,
145, 69, 2119, 285, 22, 126, 185, 56, 9, 252, 303, 2, 1],
[5, 1323, 21, 444, 121, 2120, 218, 527, 27, 9, 3350, 95,
3351], [672, 759, 110, 101, 488, 23, 50, 146, 150, 146,
235, 445, 266, 3352], [219, 52, 973, 673, 165, 206, 218,
974, 1324, 9, 89, 2, 1, 3353, 267, 1121],...]
```

Gambar4.7 *Text sequences*

Ketika *sequence* telah dibuat, Langkah berikutnya adalah menambahkan *padding*. *Padding* adalah proses untuk membuat setiap vektor kalimat memiliki panjang yang sama dengan menggunakan library `pad_sequence` dengan panjang maksimal kata adalah 100. Pada gambar 4.8 merupakan hasil dari *padding*.

```
array([[852, 967, 486, ..., 0, 0, 0],
      [1318, 757, 1319, ..., 0, 0, 0],
      [1322, 145, 69, ..., 0, 0, 0],
      ...,
      [ 632, 188, 652, ..., 0, 0, 0],
      [ 109, 345, 552, ..., 0, 0, 0],
      [ 692, 51, 10, ..., 0, 0, 0]],)
```

Gambar4.8 *Padding*

Penambahan *padding* merupakan langkah akhir sebelum dilakukan *split data*. Pada penelitian ini dilakukan perbandingan pada proses *split data* untuk membagi *dataset* menjadi data latih dan data uji, dimana terdapat 2 model *split data* yang akan dilakukan yaitu model A dan model B. Berikut merupakan inisiasi nilai pada *split data* :

- *Split data* model A = 0.2
- *Split data* model B = 0.3

Pada gambar 4.9 merupakan kode program *split data* atau membagi *dataset* menjadi data latih dan data uji.

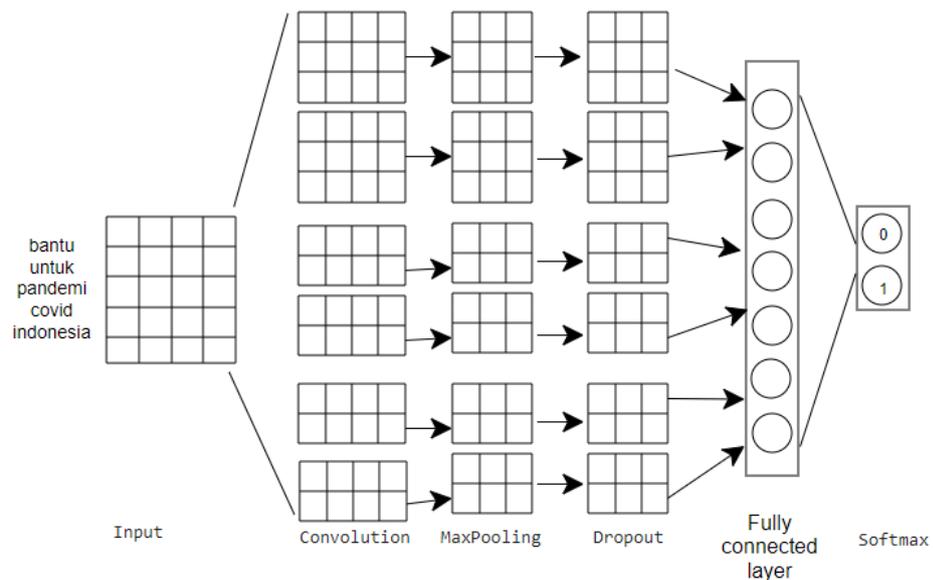
```
X_train, X_test, y_train, y_test = train_test_split(X, y,
random_state = 40, test_size= 0.2, stratify = y)
X_train, X_test, y_train, y_test = train_test_split(X, y,
random_state = 40, test_size= 0.3, stratify = y)
```

Gambar4.9 Kode Program *Split Data*

Pada Gambar 4.9 menunjukkan bahwa terdapat dua *split data* dimana pada model A dilakukan pembagian dataset berupa *test\_size* = 0,2 yang berarti dari 2.835 data dibagi menjadi data uji sebanyak 20% atau 567 data dan sisanya 80% atau 2.268 merupakan data latih. Sedangkan untuk model B dilakukan pembagian dataset berupa *test\_size* = 0,3 yang berarti dari 2.835 data dibagi menjadi data uji sebanyak 30% atau 851 data dan sisanya 70% atau 1.984 merupakan data latih. Selain itu, pada sistem juga ditambahkan *random\_state* yang merupakan angka pengacakan pemilihan atau *random number generator*(RNG) sebanyak 40 langkah.

### 4.1.3 Membangun Model Convolutional Neural Network

Pada tahapan ini, sebelum membangun model *convolutional neural network* penulis akan membuat ilustrasi arsitektur model *convolutional neural network* yang terdapat pada Gambar 4.10.



Gambar 4.10 Arsitektur model CNN

Arsitektur model *convolutional neural network* pada Gambar 4.10 tersebut yang akan digunakan untuk membangun model, dimana terdapat lapisan *embedding*, *konvolusi*, *max-pooling*, *dropout*, dan lapisan *Fully connected* yang terdapat ReLu, *dropout* ke-2, ReLu ke-2, *max-pooling* dan yang terakhir *softmax*. Pada lapisan pertama merupakan *embedding layer* sebagai input dengan Panjang vektor sebanyak 300 *embedding*, hal tersebut memungkinkan untuk model *convolutional neural network* untuk memahami setiap kata dan mengelompokkan kata yang memiliki makna mirip agar berdekatan, dimana kata yang mirip juga memiliki nilai vektor yang mirip. Selanjutnya lapisan konvolusi dengan filter sebanyak 64 dengan 8 konvolusi, lapisan *MaxPooling*, lapisan *dropout*, lapisan *Fully connected* dan lapisan *softmax*. Setelah model dibangun selanjutnya dilakukan kompilasi yang menggunakan beberapa parameter sebagai berikut :

- Parameter optimisasi “*adam*” untuk melatih model

- Sebuah parameter *loss* untuk menentukan *loss function*
- Parameter *metrics* untuk mendapatkan performa akurasi

Dari arsitektur model dan kompilasi yang dijelaskan pada kalimat diatas diimplementasikan oleh kode program pada Gambar 4.11.

```
vec size = 300
model = Sequential()
model.add(Embedding(vocab, vec size, input length =
                    max kata))
model.add(Convolution1D(64,8, activation="relu"))
model.add(MaxPooling1D(2))
model.add(Dropout(0.5))
model.add(Dense(32, activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(16, activation="relu"))
model.add(GlobalMaxPool1D())
model.add(Dense(2, activation="softmax"))
model.compile(optimizer="adam", loss =
              "categorical_crossentropy", metrics =
              ['accuracy'])
```

Gambar4.11 Kode program modelCNN

Setelah menentukan lapisan yang digunakan, tahap selanjutnya yaitu menentukan *batch size*, *verbose* dan *epoch* atau yang biasa disebut perulangan yang selanjutnya dilakukan oleh sistem pada proses *training* model. Selain terdapat data latih, juga terdapat validasi yang diimplementasikan pada Gambar 4.12. Pada sistem ini, diberikan inisiasi terhadap ketiga nilai tersebut sebagai berikut :

- *Batch size* = 32
- *Verbose* = 1
- *Epoch* = 100

```
m = model.fit(X train, y train,
              validation split=(X test, y test), epochs=100,
              batch size=32, verbose = 1)
```

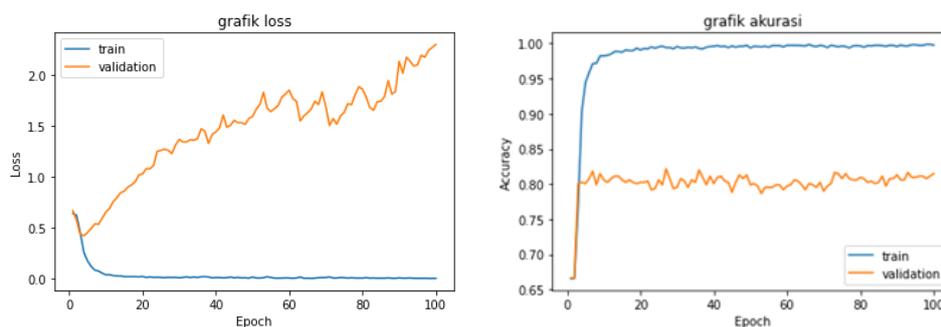
Gambar4.12 Kode program pelatihan model

Pada pelatihan model *convolutional neural network* terdapat beberapa skenario yang akan dibandingkan beserta hasil akurasi dengan menggunakan epoch 100 dimana dengan teknik tersebut tujuannya untuk menentukan model yang tepat dan cocok untuk diimplementasikan. Pada Tabel 4.3 merupakan daftar skenario yang dilakukan.

Tabel 4.3 Skenario Model *Convolutional Neural Network*

Skenario	Data Training	Optimizer	Akurasi Training
A	70%	Adam	0.9960
B	80%	Adam	0.9978
C	70%	SDG	0.9582
D	80%	SDG	0.9621
E	70%	RMSprop	0.9950
F	80%	RMSprop	0.9965

Dari skenario uji coba pada Tabel 4.3 menggunakan perbandingan dengan *data training* 70% dan 80% juga menggunakan tiga parameter *optimizer* pada model *convolutional neural network* yaitu Adam, SDG dan RMSprop mendapatkan hasil dengan akurasi tertinggi menggunakan data *training* sebanyak 80% dan menggunakan *optimizer* Adam, dari skenario B didapatkan akurasi *training* sebesar 0.9978 yang dibuktikan dengan grafik akurasi dan loss pada Gambar 4.13.



Gambar 4.13 Grafik Loss dan Akurasi

Dapat dilihat dari grafik pada Gambar 4.13 Hasil pada percobaan menggunakan model *convolutional neural network* dengan menggunakan skenario data latih sebanyak 80% pada epoch 100 terindikasi terjadi overfitting dan dari skenario tersebut menghasilkan nilai yang terdapat pada tabel 4.4.

Table 4.4 Hasil *Training* Skenario B

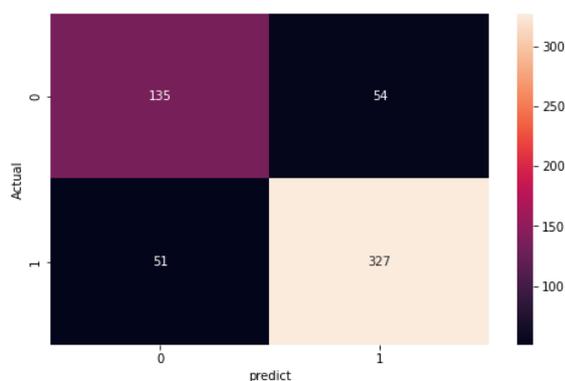
Epoch	Loss	Accuracy	<u>Val loss</u>	<u>Val accuracy</u>
100/100	0.0049	0.9978	2.3028	0.8148

Dapat dilihat pada tabel 4.4 bahwa dengan menggunakan skenario B atau data uji sebesar 80% menggunakan *optimizer* Adam pada epoch 100 didapatkan nilai *loss* 0.0049, nilai akurasi 0.9978, nilai validasi *loss* 2.3028, dan nilai validasi akurasi 0.8148 dalam waktu 7 detik. Oleh karena itu, pada penelitian ini mengimplementasikan model *convolutional neural network* dengan menggunakan skenario B.

## 1.6 Hasil Uji Coba

Pada sub bab ini menjelaskan analisis hasil pengujian sistem untuk mengetahui performa pada sistem untuk proses klasifikasi dengan data uji sebanyak

567 data yang sudah dibagi pada pembahasan sub bab 4.1.2. Untuk mengetahui nilai performa pada sistem yang sudah dibuat menggunakan *confusion matrix*. *Confusion Matrix* merupakan metode yang dipakai untuk melakukan evaluasi dan menyajikan hasil klasifikasi tentang data aktual dan data prediksi dari klasifikasi yang telah dilakukan. Pada gambar 4.14 merupakan hasil dari *confusion matrix*.



Gambar 4.14 Hasil *confusion matrix*

Dari Gambar 4.15, dapat diukur hasil evaluasi dari sistem klasifikasi menggunakan model *convolutional neural network* (CNN) sebagai berikut :

	precision	recall	f1-score	support
0	0.73	0.71	0.72	189
1	0.86	0.87	0.86	378
accuracy			0.81	567
macro avg	0.79	0.79	0.79	567
weighted avg	0.81	0.81	0.81	567

Gambar 4.15 Nilai Evaluasi

Dari hasil nilai evaluasi pada Gambar 4.15 maka peneliti melakukan percobaan pada sistem untuk menentukan klasifikasi kalimat pada tabel 4. yang mana pada tabel tersebut terdapat 15 *tweet* acak yang didapatkan dari *dataset*.

Tabel 4.5 Klasifikasi pada Sistem

Tweet	Label	
	Manual	Sistem
DANREM 172/PWY WUJUDKAN PAPIA BEBAS PANDEMI COVID-19 PERLU SINERGITAS TNI- POLRI DAN PEMDA SERTA MASYARAKAT... <a href="https://t.co/syrrZCh0TX">https://t.co/syrrZCh0TX</a>	1	1
RT @arnold5508: @ENPB_Indonesia Kinerja penanganan Pandemi COVID-19 #Indonesia status 5 Agustus 2021 - 9 Provinsi mengalami surplus dengan ...	1	1
RT @hariankompas: Melihat dinamika penanganan pandemi saat ini, ada satu urzensi untuk membenahi paradigma dan cara berkomunikasi pemerintah. ...	1	1
HMI mau kerung istana mulai 6 Agt, puncaknya 16 Agt 2021 dan alasan pemerintah gagal atasi covid-19. Org2 gagal dlm ... <a href="https://t.co/o7Km339ggU">https://t.co/o7Km339ggU</a>	0	0
Bank bjb mampu mencatatkan kinerja positif meski perekonomian tertekan akibat pandemi Covid-19. #BankBJB <a href="https://t.co/Xr2nfgG8rW">https://t.co/Xr2nfgG8rW</a>	1	1
RT @kompascom: Para ahli menyoalkan, apa strategi China selanjutnya karena saat ini sudah sangat jelas bahwa virus corona tidak akan hilang	0	0
Indonesia Diprediksi Jadi Negara Terakhir yang Keluar dari Pandemi Covid-19 <a href="https://t.co/MxiPdWBNic">https://t.co/MxiPdWBNic</a>	0	0
Keluarga mendiang Akidi Tip berencana menyumbang Rp 2 triliun untuk penanganan pandemi Covid-19 di Sumatera Selatan... <a href="https://t.co/54IjThbF6r">https://t.co/54IjThbF6r</a>	1	1
RT @Gerindra: Reses di Tengah Pandemi, Prasetyawati Bagikan Sembako, Masker, dan Buku Saku Covid-19 <a href="https://t.co/XPnm9t7tVT">https://t.co/XPnm9t7tVT</a>	1	1
Pasangan Selebritis Guntur Triyoga dan Anris Devita Jalani Bisnis Kecantikan di Musim Pandemi Covid-19	1	1

Ketua DPD RI Optimis Vaksin Nusantara Mampu Atasi Pandemi Covid-19 <a href="https://t.co/c2wEcOq8md">https://t.co/c2wEcOq8md</a>	1	1
Sebagai imbas PPKM, beberapa pengusaha warteg pada akhirnya harus menggantung status karyawan yang dirumahkan. <a href="https://t.co/KZhqk0xvNy">https://t.co/KZhqk0xvNy</a>	0	0
RT @KemenkeuRI: Akibat pandemi Covid-19, angka kemiskinan di Indonesia sempat naik, semula 9,22% pd September 2019 menjadi 10,19% pd September	0	0
Ketua DPD RI: Hikmah Dari Pandemi Covid-19 Adalah Membuka Kesadaran Terhadap Berbagai Persoalan Fundamental Bangsa <a href="https://t.co/miM9u64Rlq">https://t.co/miM9u64Rlq</a>	1	1
Bantu Warga Terdampak COVID-19, Warga Bandung Gotong Royong Siapkan Nasi Bungkus	1	1

Berdasarkan tabel 4.4 setelah dilakukannya percobaan klasifikasi dari 15 *tweet* oleh sistem didapatkan hasil yang sama dengan data yang diklasifikasi manual oleh pakar dalam bidang bahasa.

### 4.3 Pembahasan

Pembahasan yang terdapat sub bab ini berdasarkan hasil pengujian yang dilakukan pada sistem yang dibuat. Dari hasil uji coba menunjukkan bahwa sistem tersebut dapat melakukan klasifikasi sesuai *dataset* yang berisi 2.835 *tweet* yang telah didapatkan dan setelah dilakukan validasi menggunakan data uji mendapatkan nilai parameter akurasi sebesar 81%. Namun untuk nilai parameter lainnya cenderung lebih rendah dari nilai akurasi, dimana untuk nilai *precision* 73%, *recall* 71% dan *F1-Score* 72%.

Sistem klasifikasi kalimat pada penelitian ini menggunakan algoritma *convolutional neural network*, dengan data yang dipakai berasal dari media sosial. Proses pengambilan data dari media sosial yang digunakan yaitu teknik crawling dengan menggunakan API yang telah disediakan oleh media sosial twitter. Dari proses crawling didapatkan 2.835 data yang selanjutnya diberi label klasifikasi negatif atau positif. Pelabelan tersebut menghasilkan kalimat positif sebanyak 1.892 dan kalimat negatif sebanyak 943, dengan kata lain kalimat perbincangan masyarakat yang bermakna positif lebih banyak dari pada kalimat perbincangan yang bermakna negatif. Perlu kita ingat, sebagai umat beragama khususnya islam harus berpegang teguh pada Al-Quran untuk menjaga tindakan dan perilaku termasuk dalam menjaga lisan, karena manusia diberikan mulut oleh Allah untuk dapat menyampaikan segala isi hatinya dan keinginannya untuk berbuat kebaikan. Allah berfirman dalam QS Fushshilat ayat 33:

وَمَنْ أَحْسَنُ قَوْلًا مِّمَّنْ دَعَا إِلَى اللَّهِ وَعَمِلَ صَالِحًا وَقَالَ إِنَّنِي مِنَ الْمُسْلِمِينَ

*“Dan siapakah yang lebih baik perkataannya daripada orang yang menyeru kepada Allah dan mengerjakan kebajikan dan berkata, Sungguh, aku termasuk orang-orang Muslim (yang berserah diri)?”*

Isi kandungan Q.S. Fushshilat ayat 33 dalam Tafsir Departemen Agama RI (2009: 620-621) menjelaskan bahwa ayat tersebut bahwa seseorang disebut yang paling baik ketika perkataannya mengandung tiga hal, yaitu: 1) Menyeru orang lain untuk mengikuti agama tauhid, mengesakan Allah dan beriman kepada-Nya; 2) Ajakan untuk beramal shaleh, menaati perintah Allah dan menjauhi larangan dan 3) Menjadikan Islam sebagai agama utama dan memurnikan ketaatan hanya kepada Allah (Hamdani et al., 2018). Dengan kita menyampaikan hal-hal positif dan

menghindari berkata negatif berarti kita sudah melaksanakan perintah Allah dan dan menjauhi larangannya.

Berkomunikasi baik secara langsung maupun melalui media sosial yang lebih intens membuat kita lebih cepat dan membawa banyak hal baru. Namun masih banyak dari kita yang tidak menyadari setiap kalimat yang kita ucapkan atau tulis bahkan yang biasa kita baca mengandung energi yang dapat mempengaruhi kehidupan kita dan orang lain. Jadi, dengan adanya sistem ini akan sangat membantu untuk memilih kalimat yang baik dan kalimat yang buruk, sehingga komunikasi dan interaksi kita dalam kehidupan bisa mendapatkan hasil yang baik dan dapat memperoleh kemanfaatan. Allah SWT berfirman dalam surah Ibrahim ayat 26 :

وَمَثَلُ كَلِمَةٍ خَبِيثَةٍ كَشَجَرَةٍ خَبِيثَةٍ اجْتُثَّتْ مِنْ فَوْقِ الْأَرْضِ مَا لَهَا مِنْ قَرَارٍ

*“Dan perumpamaan kalimat yang buruk seperti pohon yang buruk, yang telah dicabut akar-akarnya dari permukaan bumi; tidak dapat tetap (tegak) sedikit pun.”*

Kandungan dalam Q.S. Ibrahim ayat 26 dalam Tafsir Al-Misbah menjelaskan bahwa kalimat yang buruk diperselisihkan tentang bagaimana wujudnya, yang jelas bahwa yang dimaksud dengan kalimat-kalimat yang buruk adalah contoh dari keyakinan orang-orang kafir.. Mereka sangat mudah sekali untuk diruntuhkan, tidak memiliki pijakan dan dasar yang kokoh, dan amal-amalnya juga tidak berbuah. Sehingga, kebalikan dari orang yang mukmin.

Menurut Tafsir Jalalain menjelaskan bahwa kalimat (Dan perumpamaan suatu kalimat yang tidak baik) yaitu kalimat kekafiran yang diumpamakan bagai pohon yang buruk yaitu pohon hanzhal yang buahnya sangat getir atau yang sudah

dicabut dan dibongkar sampai ke akar-akarnya dari bagian atas bumi, sehingga tidak dapat tetap sedikit pun maksudnya tidak mempunyai tempat untuk berpijak lagi, maka dari itu keadaan kalimat kekafiran tidak mempunyai tempat berpijak dan tidak pula memiliki ranting serta tidak ada keberkahannya sedikitpun. Dari kedua tafsir dari ayat tersebut sebagai pemeluk agama islam yang beriman dilarang mengucapkan kalimat buruk seperti keyakinan orang kafir karena hal tersebut tidaklah berguna dan membawa dampak negatif pada diri kita sendiri. Oleh karena itu, pada penelitian ini dibuatlah suatu klasifikasi kalimat negatif dan positif untuk membantu memilah kalimat-kalimat dan juga sebagai pelajaran bagi kita agar selalu berucap baik agar membawa dampak positif bagi seluruh umat manusia.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Pada penelitian ini menggunakan data sebanyak 2.835 *tweet* yang diklasifikasi. Klasifikasi yang telah dilakukan oleh pakar terdapat kalimat positif sebanyak 1.888 dan kalimat negatif sebanyak 947. Dari data tersebut dibagi menjadi 70% untuk data latih dan 30% untuk data uji, selain itu juga dilakukan perbandingan dengan menggunakan 80% untuk data latih dan 20% yang digunakan untuk membangun model *convolutional neural network*. Pada proses training yang telah dilakukan pada sistem menghasilkan akurasi sebesar 99,60% pada pembagian data latih sebanyak 70% dan akurasi pelatihan sebesar 99,78% pada pembagian data latih sebanyak 80%. Dari perbandingan tersebut digunakan data latih sebanyak 80% untuk proses selanjutnya karena hasil akurasi lebih tinggi.

Berdasarkan hasil training dengan 80% data latih tersebut dilakukan juga proses *testing* menggunakan data uji sebanyak 20% yang menghasilkan akurasi sebesar 81%, namun untuk nilai parameter pada proses validasi lainnya cenderung lebih rendah dari nilai akurasi, dimana untuk nilai presisi 73%, *recall* 71% dan *F1-Score* 72%. Selain itu, pada penelitian ini juga melakukan percobaan klasifikasi dengan menggunakan 15 sampel data pada sistem dimana hasil dari klasifikasi yang dilakukan oleh sistem sama dengan klasifikasi dari *dataset* yang sudah dilakukan pelabelan oleh pakar, sehingga sistem ini dapat dikatakan berjalan dengan baik karena dapat melakukan klasifikasi dengan benar.

## 5.2 Saran

Berdasarkan hasil uji coba yang telah dilakukan pada penelitian ini, diharapkan agar pada penelitian selanjutnya bisa meningkatkan akurasi pengujian model yang lebih akurat dan tidak *overfitting*. Oleh karena itu, peneliti mempunyai saran untuk pengembang melakukan penelitian pada masa yang akan datang untuk menambahkan jumlah data dan juga penambahan pada tahap *preprocessing* yang digunakan, dikarenakan jumlah *dataset* dan *preprocessing* juga dapat mempengaruhi hasil performa model. Selain itu, juga dapat menambahkan metode lain pada proses *validation* menggunakan *K-fold cross validation* agar mengurangi *overfitting* pada model dan mengembangkan metode lain pada *word embedding* seperti *word2vec*, *GloVe*, *FastText*.

## DAFTAR PUSTAKA

- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2018). Understanding of a convolutional neural network. *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017, 2018-January*, 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Annual Report Twitter. (2020). *Annual reports*.
- Brownlee, J. (2017). *Deep Learning for Natural Language Processing Develop Deep Learning Models for Natural Language in Python Acknowledgements Copyright Deep Learning for Natural Language Processing*.
- Bustami. (2014). *PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASIDATA NASABAH ASURANSI* (Vol. 8, Issue 1).
- Danukusumo, K. P. (2017). *Implementasi Deep Learning Menggunakan Convolutional Neural Network*.
- Han, J., Kamber, M., & Pei, J. (2012). Data Preprocessing. In *Data Mining* (pp. 83–124). Elsevier. <https://doi.org/10.1016/b978-0-12-381479-1.00003-4>
- Hartono, S., Sujaini, H., & Perwitasari, A. (2020). JEPIN (Jurnal Edukasi dan Penelitian Informatika). *Jurnal Edukasi Dan Penelitian Informatika*, 6.
- Hirst, G., Goldberg, Y., Williams, P., Sennrich, R., Post, M., Koehn, P., Strötgen, J., Gertz, M., Gurevych, I., Eckle-Kohler, J., Matuschek, M., Cohen, S., Veale, T., Shutova, E., Klebanov, B. B., Heinz, J., de La Higuera, C., & van Zaanen, M. (2015). *Neural Network Methods for Natural Language Processing*.
- Juwiantho, H., Setiawan, E. I., Santoso, J., Purnomo, M. H., Informasi, D. T., Tinggi, S., & Surabaya, T. (2020). SENTIMENT ANALYSIS TWITTER BAHASA INDONESIA BERBASIS WORD2VEC MENGGUNAKAN DEEP CONVOLUTIONAL NEURAL NETWORK. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(1), 181–188. <https://doi.org/10.25126/jtiik.202071758>
- Kemenkes. (2020). *KEMENKES PADK*.
- Kemenkes. (2021). *Ketahui Cara Penilaian Level Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) 1-4*.
- Kristoffer, S., & Hagelbäck, J. (2017). *Sentiment Analysis With Convolutional Neural Networks-Classifying sentiment in Swedish reviews*. [www.reco.se](http://www.reco.se)
- Kulkarni, A., & Shivananda, A. (2021). Natural Language Processing Recipes. In *Natural Language Processing Recipes*. Apress. <https://doi.org/10.1007/978-1-4842-7351-7>
- Liddy, E. D. (2001). *SURFACE SURFACE Center for Natural Language Processing School of Information Studies (iSchool) 2001 Natural Language Processing*

*Natural Language Processing* *Natural Language Processing* 1.  
<https://surface.syr.edu/cnlp>

- Listyarini, S. N., & Anggoro, D. A. (2021). Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan Convolution Neural Network (CNN). *Jurnal Pendidikan Dan Teknologi Indonesia*, 1(7), 261–268. <https://doi.org/10.52436/1.jpti.60>
- Mahendra, Y. E., Ilyas, R., & Kasyidi, F. (2020). *Klasifikasi Kalimat Ilmiah Menggunakan 1D Convolutional Neural Networks*.
- Ma'rifah, H., Prasetya Wibawa, A., & Akbar, M. I. (2020). *Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing*. 2(2), 70.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning-Based Text Classification. In *ACM Computing Surveys* (Vol. 54, Issue 3). Association for Computing Machinery. <https://doi.org/10.1145/3439726>
- Nadhifa Ayu Shafirra, & Irhamah. (2020). Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN). *Klasifikasi Sentimen Ulasan Film Indonesia Dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN)*, 9.
- Paules, C. I., Marston, H. D., & Fauci, A. S. (2020). Coronavirus Infections-More Than Just the Common Cold. In *JAMA - Journal of the American Medical Association* (Vol. 323, Issue 8, pp. 707–708). American Medical Association. <https://doi.org/10.1001/jama.2020.0757>
- Peter Jackson, & Isabelle Moulinier. (1984). *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*.
- Saad ALBAWI, Tareq Abed MOHAMMED, & Saad AL-ZAWI. (2017). *Understanding of a Convolutional Neural Network*.
- Allaam, M. R. R., & Wibowo, A. T. (2021). Klasifikasi Genus Tanaman Anggrek Menggunakan Metode Convolutional Neural Network (cnn). *eProceedings of Engineering*, 8(2).