

**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA MASA
PANDEMI COVID-19 DI MEDIA SOSIAL MENGGUNAKAN METODE NN
(NEURAL NETWORK)**

SKRIPSI

Oleh :
FEBRINA DWITA SARI
NIM. 18650001



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

**KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA
MASA PANDEMI COVID-19 DI MEDIA SOSIAL MENGGUNAKAN
METODE NN (NEURAL NETWORK)**

SKRIPSI

Oleh :
FEBRINA DWITA SARI
NIM. 18650001

**Diajukan kepada:
Fakultas Sains dan Teknologi Universitas Islam Negeri (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**

HALAMAN PERSETUJUAN

KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA
MASA PANDEMI COVID-19 DI MEDIA SOSIAL MENGGUNAKAN
METODE NN (NEURAL NETWORK)

SKRIPSI

Oleh :
FEBRINA DWITA SARI
NIM. 18650001

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 13 Juni 2022

Dosen Pembimbing I

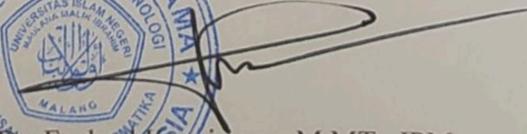
Dr. Fachrul Kurniawan, M.MT
NIP. 19771020 200912 1 001

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

HALAMAN PENGESAHAN

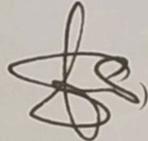
KLASIFIKASI KALIMAT PERBINCANGAN MASYARAKAT PADA MASA PANDEMI COVID-19 DI MEDIA SOSIAL MENGGUNAKAN METODE NN (NEURAL NETWORK)

SKRIPSI

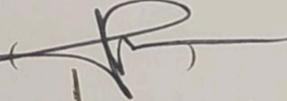
Oleh :
FEBRINA DWITA SARI
NIM. 18650001

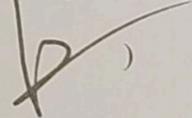
Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan Untuk Memperoleh
Gelar Sarjana Komputer (S. Kom)
Pada Tanggal: 13 Juni 2022

Susunan Dewan Penguji

Penguji Utama : Dr. M. Amin Hariyadi, M.T (
NIP. 19670018 200501 1 001

Ketua Penguji : Dr. Muhammad Faisal, M.T (
NIP. 19740510 200501 1 007

Sekretaris Penguji : Dr. Fachrul Kurniawan, M.MT., IPM (
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. 19740510 200501 1 007

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Febrina Dwita Sari

NIM : 18650001

Jurusan : Teknik Informatika

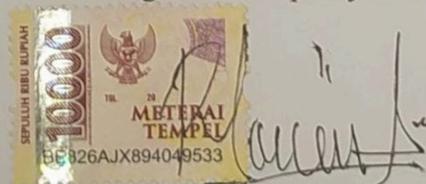
Fakultas : Sains dan Teknologi

Judul Skripsi : Klasifikasi Kalimat Perbincangan Masyarakat Pada Masa Pandemi Covid-19 Di Media Sosial Menggunakan Metode NN (Neural Network).

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, 13 Juni 2022
Yang membuat pernyataan,



Febrina Dwita Sari
NIM. 18650001

MOTTO

“Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui.”

~ QS. Al-Baqarah : 216 ~

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 Bapak Heriyanto dan Ibu Sumarniyati yang selalu tiada hentinya mendoakan dan mendukung saya hingga pada tahap ini dan saya juga mengucapkan banyak terima kasih yang sangat mendalam. Kepada keluarga besar saya yang telah mendoakan dan membantu saya dalam penyelesaian skripsi ini. Kemudian Kepada keluarga Mbak Nana dan Om Catur di Malang yang telah memotivasi dan bersedia memberikan tempat tinggal selama proses perkuliahan. Semoga kebaikan, rahmat dan perlindungan Allah SWT senantiasa tercurah kepada mereka semua. Aamiin.

Terimakasih...

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

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

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

1. Kedua orang tua saya yaitu ayah HERIYANTO dan ibu SUMARNIYATI yang telah memberikan dukungan, motivasi secara moril maupun material dan kaka kandung perempuan saya Ika Oktavianti sehingga penulis dapat menyelesaikan skripsi ini dengan tepat waktu.
2. Prof. Dr. H.M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim
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 ST., M.MT ., IPM selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang beserta jajarannya.
5. Dr. Fachrul Kurniawan ST., M.MT ., IPM dan Irwan Budi Santoso, M.Kom selaku Dosen Pembimbing, yang telah meluangkan waktu dan memberikan ilmunya kepada penulis untuk membimbing serta memberi masukan kepada penulis selama penyusunan skripsi ini hingga selesai.
6. Dr, M. Amin Hariyadi, M.T dan Dr. Muhammad Faisal, M.T 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 bermanfaat kepada saya, serta secara tidak langsung ikut terlibat dalam penyusunan skripsi ini.
8. Teman-teman Teknik Informatika Angkatan 2018 “UFO” yang telah membantu untuk bertukar pikiran dan sama-sama berjuang mengejar gelar S.Kom
9. Grup “ROAD TO SKRIPSI” beranggotakan penulis, Laila Zahrona , Yuliana Romadhoni, dan Dwi Zulva Ulinuha merupakan teman seperjuangan skripsi

yang memberikan dukungan, membantu, dan kerja sama tim dalam sepebimbingan pada proses penyelesaian skripsi ini.

10. Ari Fajariyanti selaku sahabat penulis yang telah memberikan dorongan dan motivasinya untuk menyelesaikan skripsi ini.
11. Laila Zahrona selaku teman dari awal perkuliahan sampai pada tahap penyelesaian skripsi ini yang telah menjadi teman suka dan duka selama proses perkuliahan dengan memberikan dukungan dan motivasinya kepada penulis.
12. Bintang Miftaql Huda selaku partner penulis yang telah memberikan dukungan untuk segera lulus dan semangat dalam menyelesaikan skripsi ini.
13. Seluruh teman-teman terdekat penulis Indah, Nurul, Farah, Nia, dan Berlian yang selalu mendoakan dan memberikan motivasi membangun semangat penulis dalam menyelesaikan skripsi ini.
14. Dan semua pihak yang telah ikut membantu dalam penyusunan skripsi ini hingga selesai.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan, dimana penulis berharap skripsi ini dapat memberikan manfaat kepada para pembaca termasuk penulis sendiri.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 13 Juni 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT	xv
مستخلص البحث.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	6
1.4 Batasan Masalah.....	7
1.5 Manfaat Penelitian.....	7
BAB II STUDI PUSTAKA	8
2.1 Penelitian Terkait	8
2.2 Landasan Teori	16
2.2.1 Text Mining	16
2.2.2 <i>Deep Learning</i>	17
2.2.3 <i>Natural Language Processing</i>	21
2.2.4 <i>Neural Network</i>	24
BAB III METODOLOGI PENELITIAN	37
3.1 Prosedur Penelitian.....	37
3.1.1 Identifikasi Masalah.....	38
3.1.2 Studi Literatur	38
3.1.3 Pengumpulan Data.....	38
3.1.4 Desain Sistem	38
3.1.5 Implementasi dan Pengujian Sistem	39
3.1.6 Hasil Pengujian dan Kesimpulan.....	39
3.2 Pengumpulan Data	39
3.3 Desain Sistem	41
3.4 Pre-Processing	42
3.4.1 <i>Cleaning</i>	43
3.4.2 <i>Case Folding</i>	43
3.4.3 <i>Tokenizing</i>	44
3.4.4 <i>Stemming</i>	44
3.5 <i>Word Embedding</i>	45
3.6 Alur Perhitungan <i>Neural Network</i>	47

3.7 Skenario Pengujian.....	52
BAB IV IMPLEMENTASI DAN PEMBAHASAN	54
4.1 Skenario Uji Coba.....	54
4.1.1 Tahapan Penelitian.....	55
4.2 Hasil Uji Coba.....	62
4.3 Pembahasan.....	70
BAB V KESIMPULAN DAN SARAN	76
5.1 Kesimpulan	76
5.2 Saran	77
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2. 1 Alur Proses pada Text Mining	16
Gambar 2. 2 Ilustrasi Proses pada Deep Learning	18
Gambar 2. 3 Klasifikasi Teks di Sosial Media.....	21
Gambar 2. 4 NLP dalam Artificial Intelligence	22
Gambar 2. 5 Model Sistematis pada Neuron	25
Gambar 2. 6 Arsitektur Sederhana Neural Network	27
Gambar 2. 7 Ilustrasi Jaringan Lapisan Tunggal	28
Gambar 2. 8 Ilustrasi Jaringan Saraf Banyak.....	29
Gambar 2. 9 Ilustrasi Jaringan Recurrent.....	30
Gambar 2. 10 Algoritma Fase Feed-Forward	31
Gambar 2. 11 Algoritma Fase Backpropagation.....	33
Gambar 2. 12 Grafik Fungsi Aktivasi Sigmoid Biner	35
Gambar 3. 1 Diagram Blok Metodologi Penelitian	37
Gambar 3. 2 Flowchart Desain Sistem	42
Gambar 3. 3 Flowchart Algoritma Backpropagation.....	47
Gambar 4. 1 Grafik Data Terlabeli	54
Gambar 4. 2 Hasil Akurasi Training.....	58
Gambar 4. 3 Grafik Pada Model B	59
Gambar 4. 4 Grafik Pada Model A	59
Gambar 4. 5 Grafik Pada Model D	60
Gambar 4. 6 Grafik Pada Model C	60
Gambar 4. 7 Grafik Pada Model E.....	60
Gambar 4. 8 Nilai Akurasi Testing	63
Gambar 4. 9 Hasil Confusion Matrix.....	63
Gambar 4. 10 Hasil Klasifikasi Kalimat Perbincangan di Sosial Media	65

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	12
Tabel 3. 1 Tabel Hasil Collecting Data.....	40
Tabel 3. 2 Proses Cleansing	43
Tabel 3. 3 Proses Case Folding.....	43
Tabel 3. 4 Proses Tokenizing	44
Tabel 3. 5 Proses Stemming.....	45
Tabel 3. 6 Perhitungan TF	46
Tabel 3. 7 Perhitungan IDF.....	46
Tabel 3. 8 Hasil Perhitungan TF-IDF	46
Tabel 3. 9 Data Training Hasil Pembobotan.....	48
Tabel 3. 10 Hasil Transformasi Data	49
Tabel 3. 11 Bobot Nilai Acak Input Layer ke Hidden Layer.....	49
Tabel 3. 12 Bobot Nilai Acak Hidden Layer ke Output Layer	49
Tabel 3. 13 Hasil Perubahan Bobot pada Hidden Layer	50
Tabel 3. 14 Hasil Vij Baru	51
Tabel 3. 15 Confusion Matrix	53
Tabel 3. 16 Rumus Klasifikasi Confusion Matrix	53
Tabel 4. 1 Sampel Data Hasil Pre-processing.....	56
Tabel 4. 2 Daftar Pengujian Model Neural Network.....	58
Tabel 4. 3 Skenario Uji Coba Model E.....	61
Tabel 4. 4 Hasil Klasifikasi Sampel Data	65

ABSTRAK

Dwita Sari, Febrina. 2022. *Klasifikasi Kalimat Perbincangan Masyarakat Pada Masa Pandemi Covid-19 Di Media Sosial Menggunakan Metode NN (Neural Network)*. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fachrul Kurniawan M.MT., IPM. (II) Irwan Budi Santoso, M.Kom

Kata Kunci: *Klasifikasi kalimat perbincangan, Covid-19, Media Sosial, Confusion Matrix, Neural Network.*

Covid-19 adalah sebuah penyakit yang disebabkan oleh adanya virus dan dapat menular dengan cepat kepada siapa saja. Informasi mengenai pandemi Covid-19 dan kebijakan pemerintah dalam menanganinya tersebar di berbagai sosial media Facebook. Salah satu postingan Facebook pada akun Kepmenkes RI mengenai vaksinasi di Indonesia menuai banyak komentar positif dan negatif. Komentar tersebut dapat digunakan dalam klasifikasi teks. Klasifikasi teks merupakan sebuah pengelompokan teks yang terdiri dari satu komentar positif dan negatif. Dalam penelitian ini komentar atau opini yang disampaikan oleh masyarakat tentang Covid-19 akan dilakukan klasifikasi teks. Klasifikasi teks akan dilakukan dengan mengambil data berupa komentar di sosial media, kemudian data akan dilabeli oleh ahli bahasa berupa komentar positif dan negatif. Data yang telah terlabeli akan dilakukan proses pre-processing dengan melalui empat tahapan yaitu cleaning, case folding, tokenizing, dan stemming. Kemudian data tersebut akan dilakukan pembobotan kata menggunakan TF-IDF. Setelah pembobotan kata selesai, dilakukan split data dengan membaginya menjadi dua yaitu data training dan data testing dengan perbandingan 90:10. Data training akan dilakukan pemodelan menggunakan metode *neural network*. Sehingga dari hasil pemodelan tersebut sistem dapat melakukan klasifikasi. Klasifikasi dilakukan dengan memberikan data sampel sebanyak 40 data, dimana pada data tersebut hasil klasifikasi oleh system sama dengan data terlabeli. Hal ini menunjukkan bahwa klasifikasi dengan metode neural network berhasil dilakukan dengan *accuracy* 75%.

ABSTRACT

Dwita Sari, Febrina. 2022. *Classification of Community Conversation Sentences During the Covid-19 Pandemic On Social Media Using the NN (Neural Network) Method*. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Supervisor: (I) Dr. Fachrul Kurniawan M.MT., IPM. (II) Irwan Budi Santoso, M, Kom

Keywords: *Classification of conversation sentences, Covid-19, Social Media, Confusion Matrix, Neural Network*.

Covid-19 is a disease caused by a virus and can spread quickly to anyone. Information about the Covid-19 pandemic and government policies in dealing with it is spread across various social media Facebook. One of the Facebook posts on the Indonesian Ministry of Health's account regarding vaccination in Indonesia drew many positive and negative comments. These comments can be used in text classification. Text classification is a text grouping consisting of one positive and negative comment. In this study, comments or opinions submitted by the public about Covid-19 will be classified as text. Text classification will be carried out by taking data in the form of comments on social media, then the data will be labeled by linguists in the form of positive and negative comments. The data that has been labeled will be pre-processed by going through four stages, namely cleaning, case folding, tokenizing, and stemming. Then the data will be word weighted using TF-IDF. After the word weighting is complete, the data is split by dividing it into two, namely training data and testing data with a ratio of 90:10. The training data will be modeled using the neural network method. So from the modeling results the system can perform classification. Classification is done by providing sample data as much as 40 data, where in the data the results of the classification by the system are the same as the labeled data. This shows that the classification using the neural network method has been successfully carried out with 75% accuracy.

مستخلص البحث

ويتا ساري ، فريينا. ٢٠٢٢. تصنيف جمل المحادثة المجتمعية في عهد Covid-19 على وسائل التواصل الاجتماعي باستخدام طريقة NN (Neural Network). بحث جامعي. قسم الهندسة والمعلوماتية، كلية العلوم والتكنولوجيا، الجامعة مولانا مالك إبراهيم بمالانج. المشرف: (1) دكتور فخر الكورنياوان الماجستير (2) إروان بودي سانتوسو الماجستير

الكلمات المفتاحية: تصنيف النص ، Covid-19 ، وسائل التواصل الاجتماعي ، الارتباك مصفوفة. Neural Network.

كوفيد ١٩ هو مرض ناتج عن وجود فيروس ويمكن أن ينتقل بسرعة إلى أي شخص. دخل كوفيد ١٩ لأول مرة إلى إندونيسيا في مارس ٢٠٢٠، حيث تعرض اثنان من المقيمين في المستودع لكوفيد ١٩ بعد تعاملهما مع الرعايا الأجانب اليابانيين. ثم يتزايد عدد حالات كوفيد ١٩ في إندونيسيا وينتشر في جميع أنحاء منطقة إندونيسيا. لذلك تبنت الحكومة الإندونيسية مجموعة واسعة من السياسات حول التعامل مع كوفيد ١٩ من خلال قيود الاتصال الجسدي والحجر الصحي وتطبيق 3 م في الحياة اليومية. انتشرت المعلومات حول وباء كوفيد ١٩ والسياسات الحكومية بشأن علاجه عبر فيسبوك على مواقع التواصل الاجتماعي. أحدثت إحدى تدوينات فيسبوك على الحساب المصرفي الإندونيسي حول التطعيمات الكثير من التعليقات الإيجابية والسلبية. ويمكن استخدام التعليق في تصنيف النص. تصنيف النص هو تجميع نصي من فئة أو أكثر. صنف النص في هذه العملية يمكن أن يحدد قيمة إيجابية أو سلبية على المناقشة. سيتم في هذه الدراسة إجراء التعليقات أو الآراء العامة حول كوفيد ١٩ التصنيف النصي. سيتم تصنيف النصوص من خلال جلب ملف من التعليقات على وسائل التواصل الاجتماعي، ومن ثم سيتم تصنيف البيانات من قبل اللغويين على أنها تعليقات إيجابية وسلبية. وتم معالجة البيانات من خلال أربع مراحل من التنظيف، ومعالجة القضايا، والتمييز، والاستئصال. ثم البيانات سوف تنتهي مع دمية TF-IDF بعد العمل الكامل، انقسمت البيانات إلى قسمين، بيانات التدريب وبيانات الاختبار بمقارنة بنسبة ٩٠:١٠. وستنفيذ بيانات التدريب عن طريق النمذجة باستخدام طريقة الشبكة العصبية. لذا من نتائج النمذجة يمكن للنظام القيام بالتصنيف. يتم التصنيف عن طريق إعطاء عينة من البيانات تصل إلى ٦٠ بيانات، والتي تكون نتائج تصنيف النظام مساوية للبيانات المصنفة. وهذا يشير إلى أن التصنيف بواسطة طرق الشبكة العصبية يتم بدقة ٧٥٪. صنف النص في هذه العملية يمكن أن يحدد قيمة إيجابية

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada tahun 2020, *World Health Organization* (WHO) mengkonfirmasi adanya wabah penyakit yang berasal dari Wuhan, China (*World Health Organization* (WHO), 2020). Penyakit ini yaitu virus corona atau istilah lainnya adalah *Coronavirus disease (Covid-19)*. Covid-19 adalah sebuah penyakit yang disebabkan karena adanya virus dan dapat menular dengan cepat kepada siapa saja. Virus corona pertama kali masuk ke Indonesia Maret 2020 saat terkonfirmasinya dua warga Depok yaitu ibu dan anak terpapar Covid-19 setelah menjalin dengan turis yang berasal dari Jepang(Almuttaqi, 2020).

Jumlah warga yang terpapar Covid-19 di Indonesia kemudian semakin banyak dan menyebar ke seluruh penjuru Indonesia. Maka dari itu pemerintah Indonesia menerapkan berbagai macam kebijakan dalam menangani Covid-19 mulai dari pembatasan kontak fisik (*physical distancing*), karantina, dan penerapan 3M dalam kehidupan sehari-hari. Data yang berasal dari situs resmi dalam menangani Covid-19, kasus wabah Covid-19 sampai 7 September 2021 mencatat warga yang terdampak Covid-19 adalah 42.769 jiwa dengan angka sembuh mencapai 4.023.77 jiwa. Kemudian terdapat penurunan tersebut dengan diberlakukannya aturan pembatasan kegiatan masyarakat (PPKM) dimulai dari tanggal 3 Juni – 13 September 2021, tetap mematuhi protokol kesehatan, dan melakukan vaksinasi(Krisdiyanto, 2021).

Pada masa Rasulullah pernah terjadi adanya wabah penyakit yaitu penyakit Kusta dan Lepra yang dapat menyebabkan kematian. Abdurrahman bin Auf dalam

hadistnya meriwayatkan yaitu “Apabila kamu mendengar suatu berita wabah penyakit pada suatu tempat yang dapat menular, maka janganlah kamu mendekati empat tersebut. Sebaliknya, ketika di tempat kamu terjadi wabah penyakit, maka jangan tinggalkan tempat tersebut untuk melarikan diri”. Rasulullah juga memerintahkan para sahabat untuk bersabar dalam menghadapi wabah tersebut dan selalu meminta pertolongan Allah سُبْحَانَهُ وَتَعَالَى. Dalam QS Al-Baqarah (155-157)

Allah سُبْحَانَهُ وَتَعَالَى berfirman sebagai berikut :

وَلَنَبْلُوَنَّكُمْ بِشَيْءٍ مِّنَ الْخَوْفِ وَالْجُوعِ وَنَقْصٍ مِّنَ الْأَمْوَالِ وَالْأَنْفُسِ
وَالشَّمَرَاتِ وَبَشِيرِ الصَّابِرِينَ ﴿١٥٥﴾ الَّذِينَ إِذَا أَصَابَتْهُمُ مُصِيبَةٌ قَالُوا
إِنَّا لِلَّهِ وَإِنَّا إِلَيْهِ رَاجِعُونَ ﴿١٥٦﴾ أُولَئِكَ عَلَيْهِمْ صَلَوَاتٌ مِّن رَّبِّهِمْ
وَرَحْمَةٌ وَأُولَئِكَ هُمُ الْمُهْتَدُونَ ﴿١٥٧﴾

“Dan sungguh akan Kami berikan cobaan kepadamu, dengan sedikit ketakutan, kelaparan, kekurangan harta, jiwa dan buah-buahan. Dan berikanlah berita gembira kepada orang-orang yang sabar. (156). (yaitu) orang-orang yang apabila ditimpa musibah, mereka mengucapkan: "Inna lillaahi wa innaa ilaihi raaji'uun".(157). Mereka itulah yang mendapat keberkatan yang sempurna dan rahmat dari Tuhan mereka dan mereka itulah orang-orang yang mendapat petunjuk”. (QS. Al-Baqarah :155-157)

Dalam QS. Al-Baqarah (155-157) dijelaskan bahwa Allah سُبْحَانَهُ وَتَعَالَى memberikan sebuah ujian berupa kelaparan, ketakutan, dan kekurangan harta. Bagi siapapun yang bersabar serta selalu meminta pertolongan Allah سُبْحَانَهُ وَتَعَالَى pada saat menghadapi wabah penyakit dijanjikan surga dan pahala yang besar. Covid-19 termasuk wabah penyakit yang menular dan dapat menyebabkan kematian(Hestina, 2020).

Informasi mengenai pandemi Covid-19 dan kebijakan pemerintah dalam menanganinya tersebar di berbagai sosial media. Sosial media merupakan media penyiaran yang digunakan untuk bertukar informasi, berinteraksi, dan bekerja sama dengan pengguna lainnya. Sosial media yang ramai memperbincangkan tentang Covid-19 adalah Facebook. Facebook sendiri banyak digunakan di Indonesia mulai dari remaja, orang dewasa, bahkan orang tua dari berbagai kalangan (Kementerian et al., 2020). Topik yang menjadi perbincangan masyarakat di media sosial Facebook salah satunya penanganan dan kebijakan Covid-19 oleh pemerintah yaitu vaksinasi, pemberlakuan PPKM, dan bantuan dana sosial terhadap masyarakat menengah. Kebijakan-kebijakan yang diberlakukan pemerintah menuai pro-kontra pada masyarakat tidak hanya itu pemberlakuan PPKM juga tidak luput dari komentar pedas netizen di media sosial. Berdasarkan survei nasional oleh Saiful Mujani Research and Consulting (SMRC) membuktikan bahwa masyarakat mendukung kebijakan PPKM dan masyarakat yang menolak sekitar 40-46%. Oleh karena itu masyarakat kerap kali berkomentar di sosial media Facebook.

Salah satu postingan Facebook pada akun Kepmenkes RI mengenai vaksinasi di Indonesia menuai banyak komentar positif dan negatif. Komentar tersebut dapat digunakan dalam klasifikasi teks. Klasifikasi teks merupakan sebuah pengelompokan teks yang memiliki satu atau lebih kategori. Klasifikasi teks dalam prosesnya dapat menentukan nilai positif atau negatif pada sebuah pembahasan. Dalam penelitian ini komentar atau opini yang disampaikan oleh masyarakat tentang Covid-19 akan dilakukan klasifikasi teks. Hal ini dimanfaatkan oleh lembaga-lembaga yang berkaitan untuk dijadikan evaluasi dan pertimbangan kepada pemerintah untuk menangani serta mengambil kebijakan selanjutnya

terhadap Covid-19. Selain itu, klasifikasi perbincangan masyarakat ini dapat digunakan oleh media yang menyiarkan sebuah berita dalam mengelompokkan tanggapan atau komentar di masyarakat dikategorikan berita negatif atau positif. Kemudian klasifikasi ini dapat digunakan oleh lembaga kesehatan sebagai evaluasi pelayanan kesehatan kepada masyarakat. Model klasifikasi dibagi menjadi dua yaitu model *Deep Learning* dan *Machine Learning*.

Pada buku yang ditulis oleh (Cholissodin & Soebroto, 2020) mengemukakan bahwa metode *Deep Learning* merupakan bagian cabang dari *Machine Learning* yang terdiri atas pemodelan algoritma dengan pemrosesan tingkat tinggi yang dilakukan pada data dengan menggunakan beberapa fungsi non-linear yang terstruktur membentuk lapisan-lapisan. *Deep Learning* akan memiliki hasil yang sangat baik apabila diterapkan pada *unsupervised learning*, *supervised learning*, dan *semi-supervised learning* contoh penerapannya yaitu pengenalan citra, klasifikasi teks, pengenalan suara, dan sebagainya. *Deep Learning* memiliki beberapa algoritma yang banyak digunakan seperti *Neural Network*, *Convolutional Neural Network*, dan *Recurrent Neural Network*. Sedangkan *Machine Learning* merupakan sebuah mesin yang dikembangkan untuk mempelajari data yang ada dan data yang dihasilkan untuk melakukan tugas tertentu. *Machine Learning* memiliki 4 algoritma yaitu *Support Vector Machine*, *Naïve Bayes*, *Random Forest* dan *Decision Tree*.

Mempertimbangkan beberapa penerapan model diatas maka peneliti akan melakukan penelitian menggunakan model *Deep Learning* dengan algoritma *Neural Network* (NN) dalam klasifikasi perbincangan masyarakat pada masa pandemic Covid -19. *Neural network* adalah salah satu metode pemrosesan sebuah

informasi dengan proses jaringan saraf manusia mengikuti kerja otak pada manusia. Metode ini berawal dari kesadaran atas *complex learning system* otak manusia terdiri dari sel-sel neuron saling berhubungan. Neuron dalam saraf manusia mengalami proses secara natural yang mengatur rangsangan pada *dendrit* dijalankan lalu diterjemakan dalam aktivitas axon. Kemudian untuk proses buatan, neuron memiliki tugas untuk mengatur inputan yang akan digunakan untuk menghasilkan sebuah keluaran(*ouput*)(Ananto et al., 2019). Cara kerja *neural network* yaitu dengan mengilustrasikan serta mengelompokkan beberapa bentuk yang tidak sama berasal dari sebuah populasi yang diketahui. Proses tersebut berpengaruh terhadap populasi yang terpisah sehingga dapat menentukan fungsi pembeda antar kelompok, kemudian mengklasifikasikannya menjadi sebuah objek baru dalam suatu kelompok. Berdasarkan cara kerja tersebut struktur dari *neural network* terbagi menjadi tiga lapisan yaitu input layer, hidden layer, dan output layer. Lapisan-lapisan tersebut diberikan bobot kemudian ditransformasi menjadi nilai input dan nilai output.

Metode *neural network* memiliki kelebihan yaitu melakukan pelatihan data yang dipilih serta memiliki *realtime operation* artinya dapat melakukan perhitungan dengan paralel(Sopian et al., 2019). Kelebihan selanjutnya yaitu kemampuan dalam membuat suatu pembelajaran berasal dari data yang digunakan untuk training dan memiliki *fault tolerant* tinggi(Euis Saraswati, Yuyun Umidah, 2021). Pada penelitian sebelumnya menggunakan metode *Neural network* (NN) yaitu dengan judul Penerapan Algoritma *Artificial Neural network* untuk Klasifikasi Opini Publik Terhadap Covid-19 dalam penelitiannya mendapatkan hasil *accuracy* sebesar 88,62%, *precision* sebesar 91,5%, dan *recall* 97,53%.

(Rohmawan, 2013) dengan judul *Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network* dalam penelitiannya menghasilkan persentase dengan metode Artificial Neural Network (ANN) sebesar 79,74% dibandingkan dengan metode *Decision Tree* sebesar 74,51%. (Bima et al., 2020) dengan judulnya *Analisis Sentimen Data Pada Bpjs Kesehatan Dengan Metode Backpropagation Neural Network* dalam penelitiannya mendapatkan persentase *precision* sebesar 86.97%, *recall* sebesar 87.14%, dan *accuracy* sebesar 87,14%.

Dari uraian tersebut mengenai referensi penelitian sebelumnya menggunakan metode *neural network* menunjukkan bahwa metode ini sangat menarik digunakan untuk klasifikasi text yaitu perbincangan masyarakat yang diharapkan mempunyai hasil akurasi tinggi. Maka dari itu, penelitian ini akan menggunakan metode *neural network* untuk Klasifikasi Perbincangan Masyarakat Pada Masa Pandemic Covid-19. Penelitian ini akan memberikan hasil evaluasi dengan menghitung nilai performa berupa *accuracy*, *precision*, *recall*, dan *F1-Score* dengan rumus *confusion matrix*.

1.2 Rumusan Masalah

1. Bagaimana tahapan *pre-processing* terhadap data yang digunakan?
2. Bagaimana penerapan metode *neural network* untuk klasifikasi perbincangan masyarakat pada masa pandemi Covid-19?
3. Bagaimana hasil *accuracy*, *precision*, *recal*, dan *F1-score* dalam klasifikasi perbincangan masyarakat pada masa pandemi Covid-19?

1.3 Tujuan Penelitian

1. Mengetahui hasil *pre-processing* pada data

2. Mengetahui penerapan metode *neural network* dalam mengklasifikasikan perbincangan masyarakat pada masa pandemi Covid-19
3. Membangun model *neural network* untuk klasifikasi sentimen masyarakat tentang Covid-19

1.4 Batasan Masalah

1. Data berupa komentar pada akun Facebook Kepmenkes RI berbahasa Indonesia sebanyak 1999 data.
2. Data diambil dari September 2020 – September 2021.
3. Menggunakan metode pembobotan *TF-IDF* pada tahap kedua setelah *pre-processing*.
4. Klasifikasi dilakukan dengan dua kategori yaitu positif dan negatif.
5. Tahap *pre-processing* data dibersihkan dengan cleaning yaitu menghilangkan emoticon, tanda baca, tagar, dan html untuk memenuhi proses pada text mining.

1.5 Manfaat Penelitian

1. Memberikan rekomendasi kepada yang berkepentingan untuk menjadi bahan penelitian selanjutnya dalam keilmuan dunia sains.
2. Dalam mengklasifikasi kalimat perbincangan dapat memberikan fakta yang benar tentang virus Covid-19 untuk masyarakat agar berita hoax tidak tersebar di masyarakat.
3. Penambahan pengetahuan terhadap keilmuan mengenai metode neural network.
4. Dijadikan bahan perpustakaan sebagai rujukan untuk pengembangan penelitian selanjutnya dalam bidang teknologi.

BAB II

STUDI PUSTAKA

Studi pustaka membahas penelitian sebelumnya yang akan digunakan oleh peneliti sebagai acuan dalam mempermudah penulis dalam memahami hal-hal yang berhubungan dengan penelitian dan membandingkan dengan penelitian terdahulu.

2.1 Penelitian Terkait

Penelitian terdahulu menggunakan metode *Neural network* (NN) oleh (Euis Saraswati, Yuyun Umaidah, 2021) dalam penelitiannya menggunakan media sosial Twitter. Opini masyarakat mengenai Covid-19 memiliki sentimen positif dan negative. Data yang digunakan pada penelitian ini dibagi menjadi data training 80% dan data testing 20. Skenario uji coba yang dalam penelitian ini menggunakan 1 *hidden layer* neuron 1-10 node *learning rate* yaitu 0.01-0.05. Hasil uji coba penelitian ini mendapatkan *accuracy* tinggi sebesar 0.8862 dengan *hidden layer* 10 node dan *learning rate* sebanyak 0.02. Kemudian dilakukan evaluasi yang mengukur klasifikasi menggunakan *confusion matrix* didapatkan hasil 269 data kelas negatif. Penelitian ini memiliki nilai performa *accuracy*, *precision*, *recall*, dan *F1-Score* manual dan menggunakan R Studio. Hasilnya *accuracy* sebesar 88,62%, *precision* sebesar 91,5%, dan *recall* 97,53%. Hal ini menunjukkan bahwa tergolong dalam kategori *excellent classification*, artinya proses klasifikasi yang telah dilakukan mendeteksi dengan tepat.

(Rohmawan, 2013) dalam penelitiannya dengan memberikan prediksi cara mengetahui mahasiswa dapat lulus tepat waktu dengan membandingkan dua metode diatas. Tahap awal adalah mengumpulkan data akademik seluruh mahasiswa yang kemudian diproses sampai menghasilkan data *fixed* dengan

praproses data. Data yang digunakan terbagi menjadi dua yaitu data training berupa sampel data induk dan akademik mahasiswa mulai angkatan 2006-2010 yang telah dinyatakan lulus. Data berikutnya adalah data target yang merupakan data induk dan akademik mahasiswa angkatan 2011-2012. Peneliti dalam mendapatkan data yang dibutuhkan menggunakan beberapa teknik yaitu riset perpustakaan dan riset lapangan. Berdasarkan teknik tersebut diperoleh sebanyak 600 data yang kemudian diolah kembali menggunakan tiga teknik yaitu *data validation*, *data integration and transformation*, dan *data size reduction and dicritization*. Dari hasil proses tersebut mendapatkan data yang berkualitas yaitu 553 data yang kemudian dibagi menjadi data training sebanyak 368 data dan data target sebanyak 185 data. Data tersebut diolah menggunakan metode *decision tree* dengan menggunakan *RapidMiner* yang menghasilkan tingkat *accuracy* sebesar 74.5%. Sedangkan untuk metode *Artificial Neural Network* dengan menggunakan *RapidMiner* mendapatkan hasil pengujian berupa *accuracy* sebesar 79.74%. Perbandingan hasil yang signifikan terhadap *accuracy* dari kedua metode tersebut disebabkan karena adanya perbedaan dimana metode *decision tree* bersifat sedangkan metode *artificial neural network* bersifat heuristic. Kemudian dilakukan mining yang melibatkan data dibagi menjadi dua yaitu data dengan label dan data tanpa label berupa *time series*.

(Bima et al., 2020) dalam penelitiannya mengemukakan adanya isu kenaikan iuran terkait BPJS Kesehatan yang menyebar di sosial media Twitter. Peneliti menganalisis isu tersebut untuk mendapatkan hasil tergolong opini yang bersifat negatif atau positif. Tahap awal pada penelitian ini adalah mengumpulkan komentar yang berasal dari Twitter menggunakan library python *GetOldTweets3* format. Kemudian data diklasifikasikan melalui tahap *Pre-Processing*. Selanjutnya

tahapan TF-IDF yang berfungsi agar setiap kata dijadikan sebuah vector. Setelah vector terbentuk di proses kembali menggunakan metode *Backpropagation*. Berdasarkan tahapan tersebut dilakukan pengujian data dimana data yang didapatkan sebanyak 1400 data. Data ini akan dilakukan partisi sebanyak 10 kali yang sebelumnya dibagi menjadi dua yaitu data training dan data testing dengan jumlah yang berbeda. Setelah dilakukan partisi di dapatkan hasil tertinggi pada urutan kesembilan dengan *accuracy* sebesar 87,14%. Tahap selanjutnya adalah pengujian *epoch* yang dilakukan sebanyak 5 kali dengan menggunakan learning rate sebesar 0.01 mendapatkan hasil 250. Berdasarkan hasil epoch tersebut hasil pengujian mendapatkan persentase *precision* 86.97%, *recall* sebesar 87.14%, dan *accuracy* 87,14%.

Penelitian yang dilakukan oleh (Siregar, 2018) melakukan analisis sentiment di media sosial Twitter untuk memudahkan pengguna apabila ingin mencari tahu tentang informasi tertentu yang sulit didapatkan salah satunya adalah tokoh pilkada yang mencalonkan diri. Mengenai permasalahan tersebut peneliti melakukan klasifikasi terhadap informasi tokoh pilkada dari Twitter. Tahap pertama adalah pengambilan data yang berhubungan dengan tokoh menjelang pilkada kemudian diproses melalui *text processing* dengan beberapa tahapan yaitu tokenisasi, *transform cases*, *filter token*, *stemming*, dan *stopwords*. Selanjutnya data yang telah melalui *text processing* akan menjadi data yang terstruktur barulah dilakukan pembobotan menggunakan *term frequency*. Hasil pembobotan divalidasi untuk mengestimasi keakuratan model yang dihasilkan. Operator validasi dibagi menjadi dua subproses: subproses pelatihan dan pengujian. Pada pengujian selanjutnya, peneliti menggunakan RapidMinerStudio untuk memproses metode

neural network. Ada tiga kategori kumpulan data yang diproses oleh RapidMiner: positif, negatif, dan netral. Setelah dilakukan pengujian, diperoleh hasil confusion matrix yang memberikan akurasi 66,92%, pengenalan 62,89%, dan akurasi 68,5%. Sehingga menunjukkan bahwa pengujian kurang maksimal yang disebabkan pada tahapan *preprocessing* yang didalamnya belum terdapat *wordnet* berbahasa Indonesia dimana karakteristik didalamnya terdapat kata yang berimbuhan *prefix* dan *suffix* sehingga untuk menghilangkan kata tersebut diperlukan beberapa algoritma lagi untuk mendapatkan kata dasar yang tepat.

Penelitian oleh (Hidayat et al., 2021) menganalisis terhadap kualitas perusahaan publik untuk meningkatkan reputasi perusahaan dengan menggunakan media sosial *twitter* menggunakan metode *Deep Neural Network*. ANN merupakan salah satu sistem komputasi yang dibuat dengan menirukan konsep jaringan saraf pada manusia. Jaringan syaraf ini diubah menjadi graf yang memiliki bobot dan arah. Penerapan konsep dari ANN adalah metode *deep neural network*. Metode ini dipilih karena dalam prosesnya menggunakan beberapa layer yang akan mendapatkan hasil lebih baik dari metode yang lain. Data sentiment yang digunakan dalam penelitian ini adalah *tweet* yang dituliskan masyarakat Indonesia terhadap beberapa perusahaan sebanyak 5504 data yang didapatkan melalui *preprocessing*. Tahapan dalam *preprocessing* terdiri dari *cleansing*, *case folding*, *formalization*, *steaming*, dan *tokenization*. Selanjutnya tahapan menggunakan metode DNN dimana klasifikasi yang digunakan melalui dua tahap yaitu menggunakan *training set* berfungsi untuk melatih model dan *test set* berfungsi untuk menguji model yang telah dilatih. Klasifikasi teks yang digunakan dalam penelitian ini adalah *word embedding* berfungsi untuk mengubah setiap kata

menjadi vector berisi angka. Tujuan dilakukan klasifikasi teks tersebut untuk mengenali fitur pada setiap kata. Penelitian ini menggunakan tiga kategori yaitu positif, negative, dan netral. Penelitian ini menggunakan dua scenario dengan masing-masing perbandingan 90:10 dan 80:20. Pembagian skenario pertama, data latih dan data test dilakukan dengan perbandingan 90:10. Data training yang digunakan 3512, sedangkan data testing sebanyak 390. Pembagian skenario kedua, data latih dan data test dilakukan dengan perbandingan 80:20. Sebanyak 3122 untuk data latih dan 780 untuk data test. Setelah data latih dan data test terbentuk, dilakukan analisis dengan DNN. Tahap pertama yaitu membuat *vocabulary* dan *sequence* pada setiap kalimat dalam data set. Pembuatan *vocabulary* menggunakan *library* Tensorflow. Pembuatan *sequence* sebanyak dua kali yaitu untuk data latih dan data test. Dalam *sequence* tersebut diberikan *padding post* di setiap kalimatnya untuk menyamakan jumlah token kata yang diklasifikasikan dengan DNN. Hasil eksperimen pertama dengan perbandingan 90:10 mendapatkan hasil akurasi sebesar 88,72% sedangkan perbandingan 80:20 mendapatkan hasil akurasi sebesar 82,56%. Berikut perbedaan dan persamaan penelitian terdahulu pada Tabel 2.1.

Tabel 2. 1 Penelitian Terdahulu

No	Nama	Penelitian Terdahulu	Kesamaan	Perbedaan
1.	Euis Saraswati, Yuyun Umaida, Apriade Voutama (2021)	Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19	<ul style="list-style-type: none"> - Membahas tentang Covid-19 - Menggunakan Metode Artificial Neural Network - Melalui tahapan <i>pre-processing</i> 	<ul style="list-style-type: none"> - Data yang diambil dengan collecting data pada Facebook. - Penelitian terdahulu melalui 5 tahapan pre-processing dengan perbandingan data 80:20 dan pengukuran

				akurasi menggunakan R studio. Sedangkan penelitian ini melalui 4 tahapan pre-processing dengan perbandingan data 90:10 serta pengukuran akurasi menggunakan confusion matrix menggunakan pyhon.
2.	Eko Prasetiyo Rohmawan (2021)	Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network	- Menggunakan metode Artificial Neural Network	- Menggunakan metode neural network yang menerapkan algoritma backpropagation - Penelitian terdahulu menggunakan <i>RapidMiner</i> untuk mengimplementasikan metode, sedangkan pada penelitian ini menggunakan python - Data yang digunakan penelitian ini lebih besar yaitu 3000 data sedangkan penelitian terdahulu 553 data
3.	Ricky Bima Pratama Ardika, Budhi Irawan,	Analisis Sentimen Data Pada Bpjs Kesehatan Dengan Metode	- Menggunakan algoritma backpropagation - Melalui pre-processing dan	- Data penelitian terdahulu diperoleh dari crawling twitter yang membahas

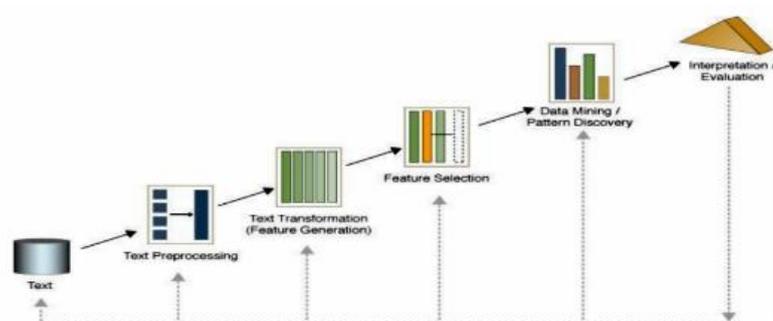
	Casi Setianingsih (2020)	Backpropagation Neural Network	pembobotan TF-IDF	<p>isu BPJS, sedangkan penelitian ini data diperoleh dengan collecting data pada Facebook.</p> <ul style="list-style-type: none"> - Penelitian terdahulu melalui 6 tahapan pre-processing, sedangkan pada penelitian ini melalui 4 tahapan - Perhitungan akurasi penelitian terdahulu dengan menguji nilai <i>epoch</i> sebanyak 5 kali, sedangkan pada penelitian ini menggunakan <i>confusion matrix</i>.
4.	Amril Mutoi Siregar (2018)	Klasifikasi Algoritma Tf Dan Neutral Network Dalam Sentimen Analisis	<ul style="list-style-type: none"> - Menggunakan neural network untuk klasifikasi - Data mealui tahapan <i>pre-processing</i> 	<ul style="list-style-type: none"> - Topik penelitian terdahulu mengenai informasi tokoh pilkada di Twitter, sedangkan penelitian ini mengenai Covid-9. - Penelitian terdahulu dengan 5 tahapan <i>pre-processing</i> sedangkan penelitian ini melalui 4 tahapan.

				<ul style="list-style-type: none"> - Pembobotan dilakukan dengan <i>term-frequency</i>, sedangkan penelitian ini dengan TF-IDF - Pengujian metode dengan <i>RapidMiner</i>, sedangkan penelitian ini dengan python.
5.	Erwin Yudi Hidayat, Raindy Wicaksana Hardiansya, Affandy (2021)	Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network	<ul style="list-style-type: none"> - Menggunakan Deep Neural Network - Melalui tahapan <i>pre-processing</i> 	<ul style="list-style-type: none"> - Penelitian terdahulu menganalisis opini perusahaan pada Twitter, sedangkan penelitian ini mengklasifikasikan kalimat perbincangan masyarakat pada Facebook - Penelitian terdahulu melalui 5 tahapan <i>pre-processing</i>, sedangkan penelitian ini 4 tahapan - Pembobotan data dilakukan dengan <i>word embedding</i>, sedangkan penelitian ini dengan TF-IDF

2.2 Landasan Teori

2.2.1 Text Mining

Text mining merupakan sebuah penemuan pengetahuan dalam *database* dengan bentuk teks. Istilah lain *text mining* adalah *knowledge discovery in textual database* (KDT). *Text mining* merupakan cabang dari data mining dimana memiliki nilai komersial jauh lebih tinggi dari data mining. Hal ini terjadi karena 80% setiap perusahaan memiliki dokumen yang berisi informasi dalam bentuk teks. *Text mining* secara umum, memiliki tiga tahapan yaitu *pre-processing*, operasi penggalian teks, dan *postprocessing*. Proses dalam *text mining* memiliki beberapa teknik dalam prosesnya yaitu teknik statistic, matematika, dan *artificial intelligence* (Kurniawan & Wibawa, 2021). Hasil *text mining* digunakan untuk memprediksi sesuatu yang akan terjadi pada sebuah masalah, memberikan informasi baru, menemukan pola yang sebelumnya tidak diketahui, dan membantu dalam mengambil sebuah keputusan. Dalam *text mining* dikenal istilah system klasifikasi teks yaitu sebuah proses pengkajian yang bertujuan untuk memisahkan dan mengelompokkan teks secara langsung sesuai dengan alur proses *text mining* (Firdaus et al., 2021). Berikut alur proses pada text mining :



Gambar 2. 1 Alur Proses pada Text Mining

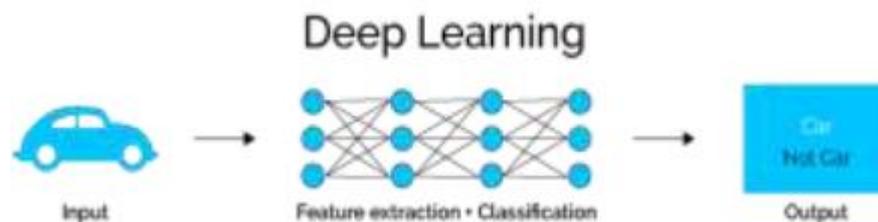
Awal mula *text mining* diperlukan sebuah pendekatan kuantitatif yang berguna sebagai analisis data yang berbentuk tekstual. Proses ini dapat mempercepat penemuan sebuah ilmu pengetahuan dengan cara meningkatkan jumlah data yang akan dianalisis. *Text mining* biasa menggunakan teknik analisis berupa *distance and similarity computing*, *dimensionality reduction*, klasifikasi, pemodelan topik dan *clustering*. *Text mining* dapat diterapkan untuk membangun sebuah penelitian yang meneliti motif pemimpin dalam sebuah perusahaan atau visi perusahaan. Salah satu penelitian dengan motif tersebut pernah dilakukan oleh J.Lee & Hong pada tahun 2013. *Text mining* berupa laporan perusahaan atau pernyataan CEO yang dibentuk menjadi data tekstual. Kemudian diselidiki dari waktu ke waktu sehingga menghasilkan pola evolusi model bisnis (Kobayashi et al., 2018).

Dalam penelitiannya (Oktasari et al., 2016) mengemukakan bahwa *text mining* merupakan sebuah teknik dalam penambangan data yang berbentuk teks. *Text mining* dalam penelitian ini untuk menganalisis sentiment pada sebuah perusahaan asuransi yaitu Prudential Indonesia di sosial media *Facebook* dengan menggunakan metode *Naïve Bayes*. Penelitian terdiri dari beberapa tahapan yaitu identifikasi data input, *pre-processing*, dan klasifikasi menggunakan metode naive bayesian. Peneliti membagi data masukan menjadi dua bagian yaitu data latih dan data uji. Hasil penelitian mencapai akurasi tinggi 95%.

2.2.2 Deep Learning

Deep learning adalah cabang pembelajaran mesin yang menggunakan jaringan saraf dalam untuk memecahkan masalah. *Deep neural network*

merupakan sebuah *neural network* dimana dalam tahapannya memiliki sejumlah hidden layer lebih dari satu (Rasywir et al., 2020). *Deep learning* merupakan sebuah metode pembelajaran dengan menirukan cara kerja system otak pada manusia yang kemudian ditirukan dan diimplementasikan pada sebuah mesin. *Deep learning* merupakan bagian dari *Machine Learning* dan *Artificial Intelligence* (AI), yang terus memahami dalam berbagai level pada abstraksi yang berbeda, umumnya menggunakan *artificial neural network* (ANN). Tantangan peneliti dalam mengimplementasikan *deep learning* adalah pengaturan nilai dalam beberapa hiperparameter contohnya adalah topologi jaringan. Topologi jaringan erat kaitannya dengan jumlah lapisan tersembunyi dan jumlah neuron tersembunyi. Menentukan jumlah lapisan tersembunyi dan neuron sangat penting sehingga berdampak signifikan pada kinerja pembelajaran yang mendalam (Choldun & Surendro, 2018). Berikut gambaran ilustrasi proses pada *Deep learning*:



Gambar 2. 2 Ilustrasi Proses pada *Deep Learning*

Teknik yang digunakan dalam *deep learning* memiliki kemiripan dengan *neural network* yaitu Restricted Boltzmann Machine (RBM) bertujuan untuk mempercepat proses pembelajaran yang menggunakan layer cukup banyak. Dengan menggunakan *deep learning* memberikan manfaat yaitu waktu yang dibutuhkan untuk training menjadi lebih singkat karena permasalahan mengenai

hilangnya gradient pada propagasi balik semakin menurun (Cholissodin & Soebroto, 2020).

Secara umum, *deep learning* terbagi menjadi tiga yaitu *Deep learning* untuk pembelajaran tanpa pengawasan atau yang dihasilkan, jaringan dalam untuk pembelajaran yang diawasi, dan jaringan dalam hibrida. Jaringan dalam untuk pembelajaran tanpa pengawasan atau yang dihasilkan adalah kategori yang menangkap korelasi tingkat tinggi dari data pengamatan atau data yang terlihat untuk analisis pola atau tujuan sintesis ketika informasi tentang label kelas target tidak tersedia. Jaringan dalam pembelajaran yang diawasi adalah kategori yang secara langsung mengklasifikasikan pola dengan menyediakan fitur. Secara umum dilakukan pada data yang tampak dengan mengkarakterisasi distribusi posterior kelas. Data label target adalah data yang selalu ada dan diawasi dalam bentuk langsung atau tidak langsung. *Hybrid deep networks* merupakan kategori yang bertujuan untuk diskriminasi yang dibantu secara signifikan dengan hasil ini muncul dari jaringan dalam yang generatif atau tanpa pengawasan. Hal ini dapat dicapai dengan meningkatkan optimasi atau regulasi jaringan. (Deng & Yu, 2013).

Secara umum arsitektur *deep learning* memiliki banyak lapisan tersembunyi yaitu operasi multi-level nonlinier. Operasi ini mentransfer setiap tingkat representasi dari input asli ke tingkat yang lebih tinggi, representasi yang lebih abstrak yang bertujuan untuk menemukan struktur yang kompleks dan unik. Salah satunya yaitu fitur-fitur bagian sudut, tepi, kontur pada sebuah gambar. Representasi abstrak ini akan dimasukkan ke lapisan klasifikasi. *Deep learning* secara keseluruhan merupakan struktur pembelajaran *end-to-end* yang

melibatkan manusia dengan parameter-parameter yang dilatih bersama(Choldun & Surendro, 2018).

Saat ini *deep learning* berkembang sangat pesat, disebabkan kemampuannya untuk mempelajari data yang digunakan untuk mendeteksi atau klasifikasi. Klasifikasi adalah analisis data yang mengkaji sekumpulan atau pengelompokan teks dan memperoleh model yang dapat digunakan kelas untuk mengelompokkan teks-teks yang masih belum diketahui(Komentar et al., 2012). Klasifikasi dibagi menjadi dua bidang yaitu klasifikasi terawasi dan tidak terawasi. Keuntungan dari klasifikasi terbimbing adalah kontrol informasi kelas berdasarkan data pelatihan dan kontrol akurasi klasifikasi. Kelemahannya adalah hasil interpretasi data terkendala, pemilihan data latih tidak jelas representatif, dan ada kelas spektral yang tidak representatif. Sedangkan kelebihan *unsupervised classification* adalah apabila terjadi kesalahan dalam operator hal tersebut dapat diminimalisir dan kelas uniq dijadikan sebagai distinct unit. Kelemahannya adalah hasil korespondensi tidak jelas terhadap kelas, kemudian setiap kelas memiliki kontrol yang terbatas, dan kelas spektral menjadi tidak konstan(Septiani et al., 2019).

Pada saat ini klasifikasi teks sangat dibutuhkan karena banyaknya jumlah produksi teks yang berkembang sangat pesat. Data teks yang berkembang pesat salah satunya dalam media sosial. Hal ini mengakibatkan perlunya analisis yang melakukan klasifikasi terhadap teks tersebut untuk mendapatkan pengetahuan atau informasi baru. Gambar dibawah ini merupakan ilustrasi dari sebuah klasifikasi teks di sosial media(Mutawalli et al., 2019).



Gambar 2. 3 Klasifikasi Teks di Sosial Media

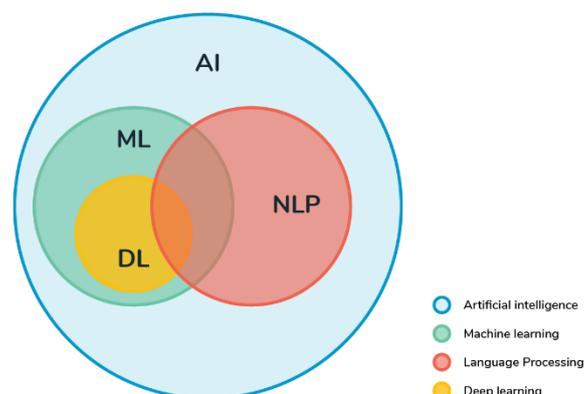
Secara umum klasifikasi teks terbagi menjadi tiga kategori, yang pertama berbasis statistic yaitu *Maximum Entropy Model*, *K-Nearest Neighbor*, *Naive Bayes*, *Support Vector Machine*, dan *Category Center Vector*. Kategori yang kedua berbasis koneksi yaitu *Artificial Neural Network*. Kategori ketiga berbasis aturan yaitu *Decision Tree* (Ridok & Latifah, 2015).

2.2.3 *Natural Language Processing*

Natural Language Processing disingkat menjadi (NLP) merupakan Sebuah pendekatan komputasi untuk menganalisis dan mewakili teks alami dalam satu atau lebih tingkat analisis linguistik untuk tujuan pemrosesan bahasa alami. Pada awalnya NLP disebut dengan *natural language understanding* yang memiliki tugas yaitu memparafrase teks masukan, Menerjemahkan teks ke dalam bahasa lain, menjawab pertanyaan tentang isi teks, dan menarik kesimpulan dari teks (Grosz, 1982).

NLP berfungsi untuk membuat interaksi manusia dan computer menjadi lebih efisien misalnya sebuah mesin yang mempelajari makna Bahasa

manusia yang kemudian diproses dan menghasilkan sebuah output. NLP sangatlah penting digunakan dimasa yang akan datang karena akan mempermudah kita adalah membangun sebuah model dan proses yang melibatkan adanya informasi sebagai input dalam bentuk teks, gambar, dan suara yang kemudian dimanipulasi kembali sesuai algoritma yang akan digunakan dalam computer (Aditya Jain, Gandhar Kulkarn, 2018). Penggunaan NLP memiliki tujuan yaitu untuk merancang dan membangun aplikasi yang dapat memfasilitasi interaksi manusia dengan komputer menggunakan bahasa alami. Bahasa yang telah diterima oleh computer nantinya akan diproses dan dipahami agar maksud dari manusia dapat dipahami oleh computer (Sartini, 2020). Berikut letak NLP di lingkup *artificial intelligence*:



Gambar 2. 4 NLP dalam *Artificial Intelligence*

Natural language processing suatu proses yang sangat penting dalam membangun model dan proses yang membutuhkan beberapa informasi sebagai *input* dalam bentuk yang beragam misalnya suara, gambar, teks maupun ketiganya. Informasi tersebut dapat dijadikan penelitian dengan melakukan manipulasi sesuai algoritma yang ada pada computer. Sehingga inputan tersebut dapat menghasilkan output dengan *natural language processing*. Salah satu

algoritma yang dikembangkan untuk meningkatkan efisiensi NLP dalam bentuk teks adalah Algoritma Neural Network (NN), Named Object Recognition Model, User Interest Graph Model, word embedding, Pattern-based algorithm using automatic text summarization, Frase extraction based on fuzzy inference rule features (Aditya Jain, Gandhar Kulkarn, 2018).

Kegunaan dari *natural language processing* yaitu digunakan sebagai pelacak komentar masyarakat di media sosial, berita, dan mendapatkan informasi yang telah dibicarakan user di setiap platform. Secara umum NLP memiliki tugas dasar yaitu dalam menentukan topik dan pemodelan, mengkonversi data menjadi kalimat begitupun sebaliknya, penerjemah mesin, analisi sentiment, dan pengkategorian atau klasifikasi teks. NLP sangat baik digunakan untuk klasifikasi teks dengan catatan hasil akurasi yang didapatkan harus mempunyai nilai yang tinggi untuk menghindari adanya salah penafsiran sehingga menyebabkan adanya kesalahan tindakan yang akan dilakukan selanjutnya. Pengolahan teks dengan NLP akan dimulai dengan analisis morfologi yang dilakukan dengan cara mengkonversi teks atau dokumen untuk mendapatkan beberapa macam morfologi kata yang terlibat (Maitri & Sutopo, 2019).

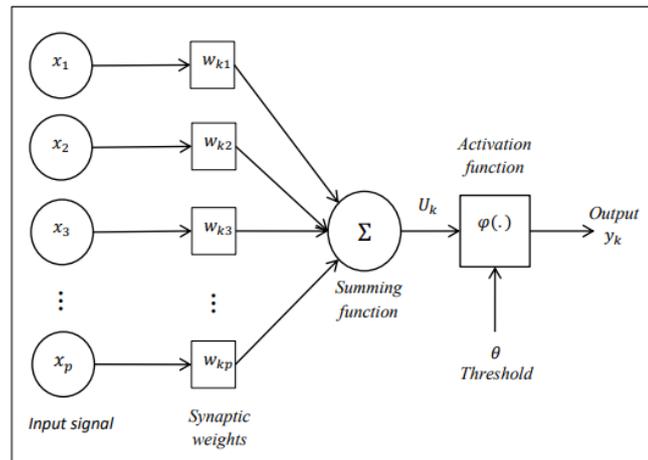
Penelitian yang dilakukan oleh (Munasatya & Novianto, 2020) melakukan analisis sentimen dengan menggunakan NLP untuk mengolah data agar mudah dibaca oleh computer. Proses pengolahan ini dikenal dengan istilah *pre-processing*. Sebelum melakukan *pre-processing* peneliti terlebih dahulu menyiapkan data yang diambil dengan cara *crawling* data *tweet*. Kemudian data tersebut akan di *pre-processing* agar data menjadi bersih. Hal ini karena data

yang diperoleh dari *crawling* tadi belum bersih masih bercampur dengan tanda baca, tagar, emoticon dsb. Setelah data bersih barulah dilakukan pelabelan data dengan tiga kategori yaitu positif, negative, dan netral *lexicon based* yang melakukan perhitungan polaritas dengan kamus kata. Selanjutnya, menggunakan metode multi layer untuk proses klasifikasi untuk mengetahui nilai akurasi *tweet* tentang Presiden Jokowi. Penggunaan NLP dalam penelitian ini akan mendapatkan hasil secara otomatis berupa teks dan juga label dengan akurasi yang cukup tinggi sebesar 93,26%, kemudian sebesar 95,36% nilai *F-measure*, dan *recall* sebesar 94,21%.

2.2.4 Neural Network

Jaringan saraf tiruan adalah ilmu komputer yang mengikuti jaringan saraf otak manusia, di mana arsitektur dan pengoperasian sistem memberikan rangsangan. B. Melakukan aktivitas untuk menghasilkan output dari tindakan yang akan dilakukan. Jaringan saraf tiruan pertama kali dikembangkan oleh para ilmuwan oleh Warren Sturgis dan Walter Pitts pada tahun 1943. Menurutnya, *neural network* adalah sebuah pola yang memproses informasi berdasarkan sistem saraf pada manusia. Secara sederhana, sistem pemrosesan informasi jaringan saraf terstruktur terdiri dari sejumlah proses yang saling berhubungan dalam bentuk neuron yang bekerja untuk menemukan solusi dari suatu masalah. Neural network merupakan algoritma yang tepat untuk klasifikasi dalam data mining, text dan image mining. Jaringan saraf juga digunakan untuk pembelajaran mesin yang secara konseptual meniru sistem otak manusia melalui perhitungan neuron yang saling berhubungan(Siregar, 2018).

Berdasarkan karakteristik neuron, prinsip jaringan saraf tiruan memiliki nilai input, dan setiap input dikalikan dengan faktor pembobotan. Kemudian jumlahkan input yang sudah dibobot untuk menentukan tingkat aktivasi neuron. Berikut ini adalah deskripsi sistematis dari proses yang diuraikan dalam Gambar 2.5.



Gambar 2. 5 Model Sistematis pada Neuron

Gambar 2.5 adalah sebuah model matematika non-linier dengan beberapa sinyal input yang disimbolkan $x_1, x_2, x_3, \dots, x_p$ dan $w_{k1}, w_{k2}, w_{k3}, \dots, w_{kp}$ adalah bobot dari *neuron* k ; terdapat adalah *linear combiner* output yang menjumlahkan bobot dan input $w_{kp}x_p$. Fungsi aktivasi yang disimbolkan dengan $\varphi(\cdot)$, adalah threshold atau nilai batasan dan adalah sinyal output dari *neuron*. Penggunaan threshold nantinya memberikan pengaruh terhadap output U_k dari *linear combiner*. Dalam metode *neural network*, setiap sel saraf masing-masing memiliki tiga bagian yaitu fungsi penjumlahan, fungsi aktivasi, dan fungsi keluaran. Hal ini dapat digambarkan dalam persamaan sebuah *neuron* k sebagai berikut :

$$U_k = \sum_{j=1}^p W_{kj} X_p \quad (2.1)$$

$$Y_k = \varphi(U_k + \theta_k) \quad (2.2)$$

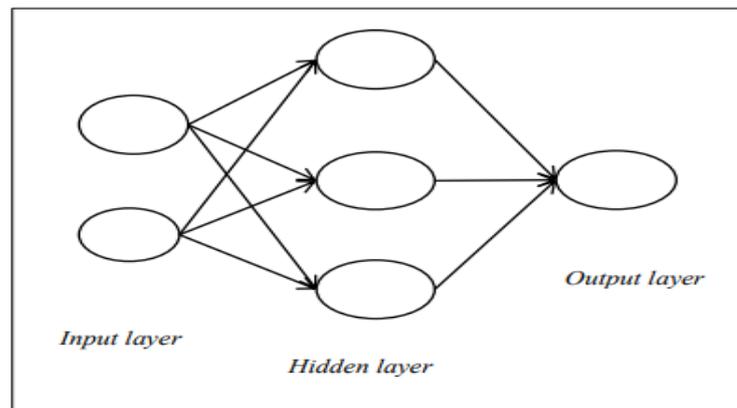
Berdasarkan persamaan diatas, dapat didefinisikan $w_{k j}$ adalah nilai antara layer *input* dan *output* ke pj dengan ($p = 1,2,3,\dots, n$ dan $j = 0,1,2,3, \dots, n$). Nilai x_p yaitu *inputan* ke ($p = 1,2,3, \dots, n$). Setiap metode pasti memiliki kelebihan dan kekurangan. Berikut kekurangan dari *neural network* menurut (Hermawan, 2018):

1. Memiliki kemampuan mengolah data dalam kondisi yang tidak memiliki kepastian dikarenakan adanya proses abstraksi , generalisasi, dan ekstraksi pada struktur datanya.
2. Dapat mempresentasikannya secara sederhana sehingga dapat menciptakan representasi *self organizing* .
3. Memberikan toleransi terhadap adanya kesalahan pada data.
4. Dapat memproses data secara efisien disebabkan memiliki system yang parallel, sehingga dapat mempersingkat waktu dalam pengoperasiannya.

Kelemahan *neural network* menurut (Sutojo, dkk., 2011):

1. Metode *neural network* tidak dapat berjalan efektif jika digunakan dalam pengoperasian numerik dengan ketelitian yang tinggi .
2. Jika menggunakan operasi logika dan algoritma aritmatik menjadi tidak efisien.
3. Memerlukan adanya data training, dikarenakan waktu pengoperasian akan sangat lama jika data yang digunakan banyak.

Secara umum arsitektur *neural network* memiliki tiga *layer* yaitu *input*, *hidden*, dan *output layer*. Lapisan tersebut memiliki *node* dan *neuron* dengan jumlah yang berbeda. Berikut gambar 2.6 merupakan arsitektur *neural network* sederhana.



Gambar 2. 6 Arsitektur Sederhana Neural Network

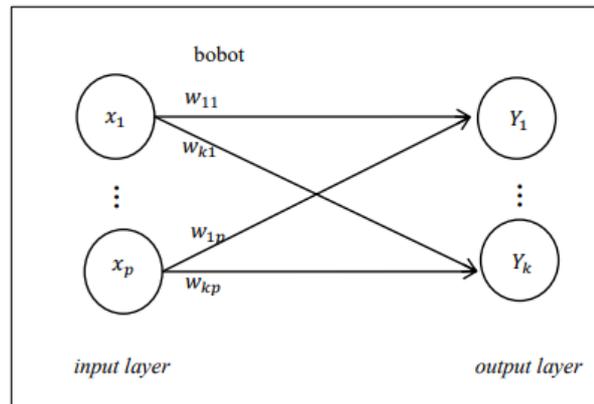
Berdasarkan gambar 2.6, berikut penjelasan setiap arsitektur NN :

- a. Lapisan *input* merupakan lapisan yang terdiri dari beberapa *neuron* yang menerima sinyal dari luar kemudian dilanjutkan ke dalam *neuron* lain yang terdapat pada jaringan berdasarkan cara kerja saraf.
- b. Lapisan *hidden layer* merupakan lapisan yang dibuat untuk menirukan sel saraf dan terkoneksi pada saraf asli yang berguna untuk meningkatkan kemampuan jaringan itu sendiri.
- c. Lapisan *output layer* merupakan lapisan yang digunakan untuk jalan hasil keluaran dalam proses jaringan berupa sel saraf tiruan motoric pada jaringan saraf asli.

Penelitian yang dilakukan oleh (Fausett, 2004) menyatakan bahwa *neural network* dalam berbagai aplikasi memiliki beberapa jenis arsitektur yang biasa digunakan sebagai berikut :

1. Jaringan Lapisan Tunggal

Jaringan ini memiliki satu lapisan dari nilai koneksi. Lapisan *input* akan terhubung langsung ke lapisan *output* akan tetapi tidak terhubung dengan unit berbeda dalam lapisan yang sama. Berikut gambaran jaringan lapisan tunggal.

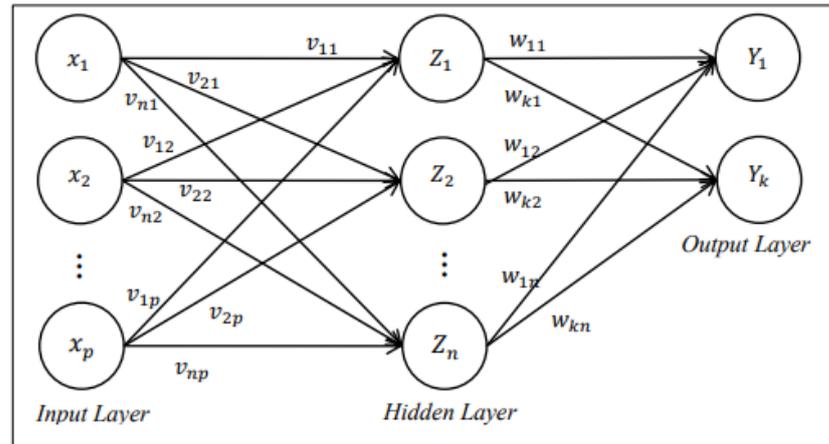


Gambar 2. 7 Ilustrasi Jaringan Lapisan Tunggal

Dapat dilihat dalam gambar 2.7 merupakan arsitektur jaringan lapisan tunggal n dengan lapisan *input* $(x_1, x_2, x_3, \dots, x_p)$ dan k sebagai *output* $(y_1, y_2, y_3, \dots, y_k)$.

2. Jaringan Lapisan Banyak

Jaringan ini adalah jaringan yang terdiri dari satu atau lebih lapisan dari simpul antara unit *input* dan unit *output*. Jaringan ini memiliki sebuah lapisan diantara dua tingkat yang saling berdekatan dari unit *input*, *hidden layer*, dan *output*. Berikut gambaran jaringan lapisan banyak.

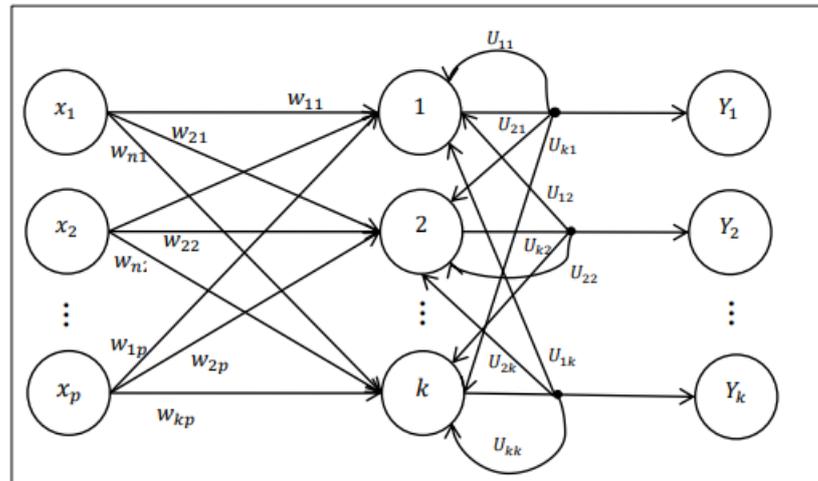


Gambar 2. 8 Ilustrasi Jaringan Saraf Banyak

Berdasarkan gambar 2.8 merupakan jaringan lapisan banyak dimulai p adalah lapisan *input* ($x_1, x_2, x_3, \dots, x_p$), selanjutnya lapisan *hidden* memiliki n unit ($Z_1, Z_2, Z_3, \dots, Z_n$), kemudian terakhir terdiri dari k yang merupakan lapisan *output* ($y_1, y_2, y_3, \dots, y_k$).

3. Recurrent Network

Jaringan ini memiliki kemiripan dengan lapisan tunggal dan lapisan banyak. Akan tetapi memiliki *output* yang memberikan sinyal pada *input*. *Recurrent network* memiliki karakteristik yang sederhana karena hanya terdiri dari satu *looping* umpan balik. Berikut gambar 2.9 merupakan ilustrasi jaringan *reccurent network*



Gambar 2. 9 Ilustrasi Jaringan Recurrent

- Algoritma Backpropagation

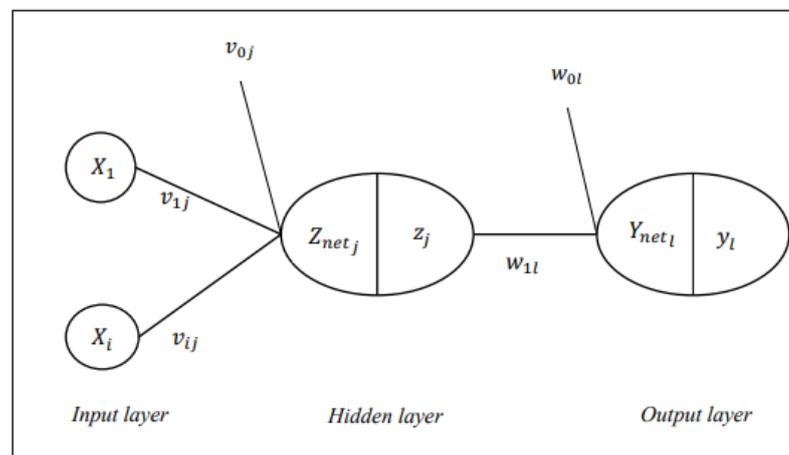
Algoritma ini merupakan algoritma dalam *neural network* yang berfungsi untuk menemukan sebuah fungsi yang bergerak berulang-ulang sehingga mendapatkan fungsi minimal dan maksimal. Dalam penelitian ini menggunakan algoritma *backpropagation* karena dalam beberapa pengaplikasiannya terdapat proses training dengan koneksi yang sederhana yaitu jika *output* yang dihasilkan memberikan hasil maka bobot akan dikoreksi kembali agar *error* dapat di minimalisir. Kemudian *neural network* memberikan tanggapan yang diharapkan mendekati nilai yang tepat sehingga algoritma ini mampu untuk memperbaiki bobot dalam lapisan *hidden layer* (Kumar Abhishek, Anshul Khairwa, Tej Pratap, 2012).

(Fausett, 2004) menyatakan bahwa algoritma *backpropagation* pada saat melakukan *training* dalam suatu jaringan memiliki tiga tahapan dimana ketiga tahapan atau fase tersebut dilakukan berulang sampai penghentian terpenuhi. Dalam kondisi tersebut jumlah iterasi akan melebihi jumlah

maksimum yang ditentukan. Berikut tahapan algoritma *backpropagation* secara rinci sebagai berikut :

1. Fase *Feed-Forward*

Dalam fase ini setiap *input* akan dihitung maju menuju lapisan *hidden* sampai lapisan *output* yang menggunakan fungsi aktivasi yang telah ditentukan. Gambar 2.10 merupakan ilustrasi algoritma *backpropagation* pada fase *feed-forward* :



Gambar 2. 10 Algoritma Fase Feed-Forward

- Pada setiap layer *input* yaitu (x_i $i = 1,2,3, \dots, p$) akan menerima sinyal *input* x_i yang kemudian diteruskan ke dalam lapisan *hidden*.
- Setiap lapisan *hidden* yaitu (Z_j $j = 1,2,3, \dots, n$) dijumlahkan dengan bobot sinyal *input* dengan persamaan :

$$Z_{net\ j} = v_{0j} + \sum_{i=1}^p x_i v_{ji} \quad (2.3)$$

Keterangan :

$Z_{net\ j}$: merupakan nilai output pada unit Z_j

x_i : merupakan nilai *input* dalam unit i

v_{ij} : merupakan bobot yang terdapat antara *input* dalam unit i dengan lapisan j

v_{0j} : merupakan bobot bias *input* unit i dan lapisan j

Langkah selanjutnya adalah mengaplikasikan fungsi aktivasi yang berfungsi menghitung sinyal. Kemudian sinyal tersebut dikirim ke semua unit yang ada di lapisan *output*.

$$z_j = f(Z_{net j}) = \frac{1}{1 + \exp^{-Z_{net j}}} \quad (2.4)$$

Keterangan :

z_j : merupakan nilai aktivasi berasal dari unit Z_j

c. Langkah selanjutnya setiap unit dari lapisan *output* ($Y_l, l = 1, 2, 3, \dots, k$) menjumlahkan bobot *output* yang didapatkan dari lapisan *hidden*.

$$Y_{net l} = w_{0k} + \sum_{j=1}^n Z_j w_{kj} \quad (2.5)$$

Keterangan :

$Y_{net l}$: merupakan *input* sebagai unit Y_l

w_{0k} : merupakan nilai bobot dari sambungan yang terdapat dalam bias sebagai unit Y_l

w_{kj} : merupakan nilai bobot dari sambungan Z_{ij} ke unit dalam Y_l

Kemudian dikalikan hasil $Y_{net l}$ dengan fungsi aktivasi agar menghasilkan *output*

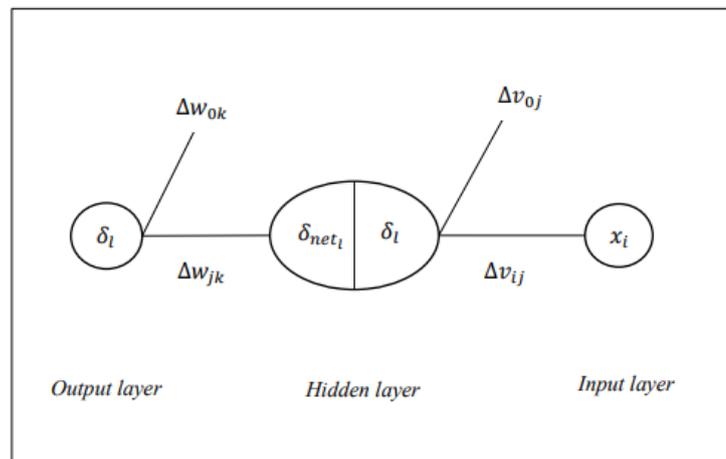
$$y_l = f(Y_{net l}) = \frac{1}{1 + (\exp^{-Y_{net l}})} \quad (2.6)$$

Keterangan :

y_l : merupakan nilai *output*

2. Fase *Backpropagation*

Pada fase ini jika terjadi kesalahan dimana terjadi selisih antara *output* dengan target yang ditentukan maka dipropagasi mundur mulai garis yang terhubung langsung dengan unit lapisan *output*. Gambar 2.11 merupakan ilustrasi fase *backpropagation*.



Gambar 2. 11 Algoritma Fase Backpropagation

- a. Unit *output* Y_l menerima gambaran atau pola hasil *training input* kemudian dilakukan perhitungan nilai *error* yang berasal dari lapisan *output*.

$$\delta_l = (t_l - y_l) f'(Y_{net_l}) \quad (2.7)$$

Keterangan :

δ_l : kesalahan nilai aktivasi dari lapisan *output*

t_l : nilai target data

Penghitungan bobot baru untuk memperbaharui w_{jk}

$$\Delta w_{jk} = \alpha \delta_l z_j \quad (2.8)$$

Keterangan :

Δw_{jk} : merupakan selisih antara $w_{kj}(t)$ dan $w_{kj}(t + 1)$

α = merupakan nilai learning rate

z_j = merupakan nilai aktivasi pada unit Z_j

b. Menjumlahkan delta *input* pada setiap lapisan hidden

$$\delta_{netl} = \sum_{l=1}^k \delta_l w_{kj} \quad (2.9)$$

Keterangan:

δ_{netl} : merupakan nilai untuk menghitung kesalahan dilapisan *hidden*

w_{kj} : merupakan bobot antara z_{ij} ke Y_l

Hasilnya dikalikan dengan turunan yang berasal dari fungsi aktivasi yang digunakan menghitung informasi *error*.

$$\delta_l = \delta_{netl} f'(z_{netj}) = \delta_{netl} z_j (1 - z_j) \quad (2.10)$$

Keterangan :

δ_l : merupakan nilai aktivasi dari kesalahan lapisan *hidden*

Langkah selanjutnya adalah menghitung koreksi bobot sebagai pembaharuan v_{ij}

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (2.11)$$

Keterangan :

Δv_{ij} : nilai selisih antara $v_{ij}(t)$ dan $v_{ij}(t + 1)$

x_i : adalah unit ke-I dalam lapisan *input*

Kemudian perhitungan koreksi bias berfungsi untuk pembaharuan v_{0j} .

$$\Delta v_{0j} = \alpha \delta_j \quad (2.12)$$

Kerangan :

Δv_{0j} : merupakan koreksi dari nilai bias.

3. Fase Penyesuaian Bobot

Fase ini melakukan adanya modifikasi bobot yang digunakan agar terjadi penurunan kesalahan.

- a. Unit pada setiap *output* Y_l merubah bobot dan biasanya

$$w_{jl}(\text{baru}) = w_{jl}(\text{lama}) + \Delta w_{jl} \quad (2.13)$$

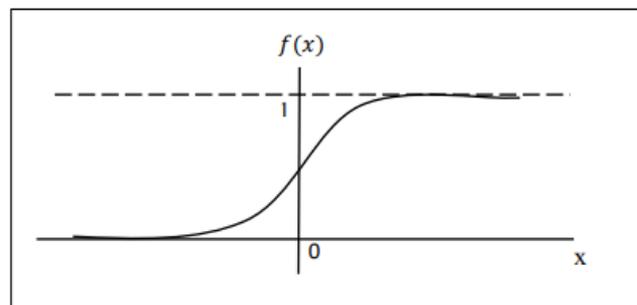
- b. Unit yang tersembunyi Z_j merubah bobot dan biasanya

$$v_{jl}(\text{baru}) = v_{jl}(\text{lama}) + \Delta v_{jl} \quad (2.14)$$

- Fungsi Aktivasi

(Puspitaningrum, 2006) menyatakan bahwa dalam metode *neural network* fungsi aktivasi memiliki peran sebagai sinyal yang menentukan *output* yang diteruskan ke beberapa *neuron*. Fungsi aktivasi inilah yang nantinya akan menentukan besar kecilnya bobot. Fungsi *sigmoid biner* apabila diterapkan dalam *neural network* akan sangat baik jika pembelajarannya dengan algoritma *backpropagation*. Fungsi *sigmoid biner* mempunyai *range* baik jika digunakan pada sebuah jaringan dengan hasil *output* 0 sampai 1 yang diilustrasikan secara sistematis pada gambar 2.12(Fausett, 2004)

$$y = f(x) = \frac{1}{1 + e^{-x}} \quad (2.15)$$



Gambar 2. 12 Grafik Fungsi Aktivasi Sigmoid Biner

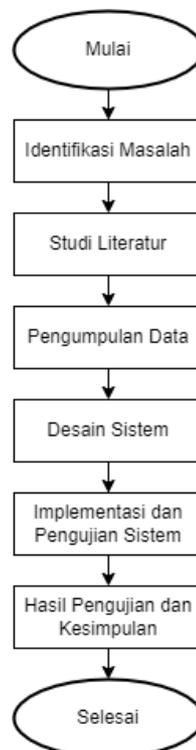
Fungsi aktivasi dalam *backpropagation* memiliki karakter salah satunya yaitu *continue* artinya jaringan Mudah diturunkan dan tidak monoton. Fungsi yang memenuhi kriteria tersebut adalah fungsi sigmoid biner. Hal ini dikarenakan fungsi ini dapat mengurangi adanya nilai input yang berlebihan dan membuat jaringan lebih kuat untuk menghasilkan output. (Fausett, 2004).

BAB III

METODOLOGI PENELITIAN

3.1 Prosedur Penelitian

Prosedur penelitian merupakan rangkaian kegiatan yang dilakukan dalam suatu survei untuk mendapatkan gambaran yang terstruktur dan sistematis. Secara umum, penelitian dapat dibagi menjadi dua pendekatan: kuantitatif dan kualitatif. Penelitian ini biasanya menggunakan pendekatan kuantitatif, dengan menggunakan sampel yang didukung oleh alat survey atau pengukuran. Penelitian ini termasuk dalam analisis data kuantitatif yang bertujuan untuk menguji hipotesis yang diperoleh. Gambar 3.1 di bawah ini menunjukkan proses survei dalam format diagram blok sehingga survei dapat berjalan lancar dan terstruktur serta mencapai tujuan yang telah ditetapkan.



Gambar 3. 1 Diagram Blok Metodologi Penelitian

Berdasarkan gambar 3.1, diagram blok merupakan tahapan-tahapan yang akan dilakukan pada penelitian ini agar selesai dengan terstruktur. Berikut penjelasan setiap blok diagram.

3.1.1 Identifikasi Masalah

Tahap ini adalah suatu proses yang sangat penting dalam penelitian dimana berisi identifikasi masalah yang diangkat dari kejadian sekitar secara nyata. Identifikasi masalah memuat rincian bagaimana masalah tersebut dapat terjadi.

3.1.2 Studi Literatur

Tahap ini adalah rangkaian kegiatan kedua dalam penelitian ini yang berhubungan dengan pengumpulan data pustaka sebagai referensi baca oleh peneliti agar dapat memahami konsep secara mendalam. Biasanya studi literatur berasal dari penelitian terdahulu misalnya jurnal, buku, dan majalah yang memiliki keterkaitan dengan tujuan dan masalah penelitian.

3.1.3 Pengumpulan Data

Tahap ini adalah kegiatan mencari data di lapangan dengan tujuan mendapatkan data yang baik berupa teks, angka, tabel dll. Data ini akan diproses oleh peneliti agar memperoleh hasil dan menjawab masalah dari penelitian.

3.1.4 Desain Sistem

Tahap ini adalah sebuah perancangan atau ilustrasi yang dapat berupa sketsa dari beberapa kumpulan elemen kemudian disatukan membentuk suatu gambaran desain yang terstruktur dan menjadi acuan peneliti dalam menyelesaikan penelitian.

3.1.5 Implementasi dan Pengujian Sistem

Tahap ini adalah tahapan dimana data yang telah di kumpulkan dan rancangan system yang telah dibuat akan di implemenaskan untuk menghasilkan sebuah system yang baik. Kemudian untuk mengetahui system layak atau tidak dilakukan pengujian terlebih dahulu.

3.1.6 Hasil Pengujian dan Kesimpulan

Hasil pengujian dan kesimpulan adalah rangkaian tahapan terakhir dalam penelitian ini berisikan penjelasan dari system yang telah dibuat.

3.2 Pengumpulan Data

Secara umum data terbagi dua yaitu data primer dan data sekunder. Dalam penelitian ini data yang digunakan adalah data primer. Data primer merupakan data yang diperoleh secara langsung istilahnya adalah data mentah. Data primer didapatkan dengan berbagai cara misalnya dengan survei, wawancara, dan pengumpulan data lainnya. Dalam penelitian ini pengumpulan data yang digunakan dengan cara collecting data dimana data yang diperoleh berasal dari komentar di media sosial facebook yang di ambil dari akun Kepmenkes RI. Data yang diambil berupa kritikan dan saran dari masyarakat terhadap kebijakan-kebijakan yang dikeluarkan pemerintah dalam menangani Covid-19. Salah satu contoh kebijakan yang dikeluarkan pemerintah yaitu melakukan *social distancing*, mentaati 3M. vaksinasi, dan pemberlakuan PPKM. Komentar dan saran yang diambil pada postingan tersebut adalah postingan pada tahun 2020 dan 2021. Data diperoleh dengan cara manual *copy paste* hal ini disebabkan API pada Facebook tidak dapat

diakses seperti dahulu sehingga untuk mengambil data secara crawling tidak memungkinkan.

Selanjutnya data dilakukan pelabelan yang dilakukan oleh ahli bahasa yang telah menguasai pembahasan seputar Covid-19. Pelabelan dilakukan dengan memberikan 10 data sampel secara acak yang kemudian akan dianalisis oleh ahli Bahasa untuk dimasukkan kedalam kategori positif atau negatif. Berdasarkan hasil analisis tersebut, peneliti diberikan formula sebagai dijadikan acuan untuk melakukan pelabelan selanjutnya pada semua data. Berikut ciri-ciri kalimat negatif dan positif dari ahli Bahasa untuk acuan melabeli semua data.

1. Kalimat positif : kalimat mengandung unsur doa, mengandung unsur optimisme, berpikir logis, kalimat kritis berdasarkan fakta, terdapat unsur berserah kepada tuhan, memotifasi untuk maju kearah yang lebih baik.
2. Kalimat negatif : terdapat unsur mengejek, terdapat unsur prasangka buruk, terdapat unsur menghasut, tidak percaya kepada pemerintah, terdapat unsur sindiran, terdapat unsur kalimat yang pesimis, terdapat unsur kalimat yang menuduh pihak lain, terdapat unsur kata yang menunjukkan ketidakpercayaan.

Berdasarkan ciri-ciri kalimat positif dan negatif dari ahli bahasa, berikut 10 sampel data acak atau komentar yang telah di labeli oleh ahli bahasa pada tabel 3.1.

Tabel 3. 1 Tabel Hasil *Collecting Data*

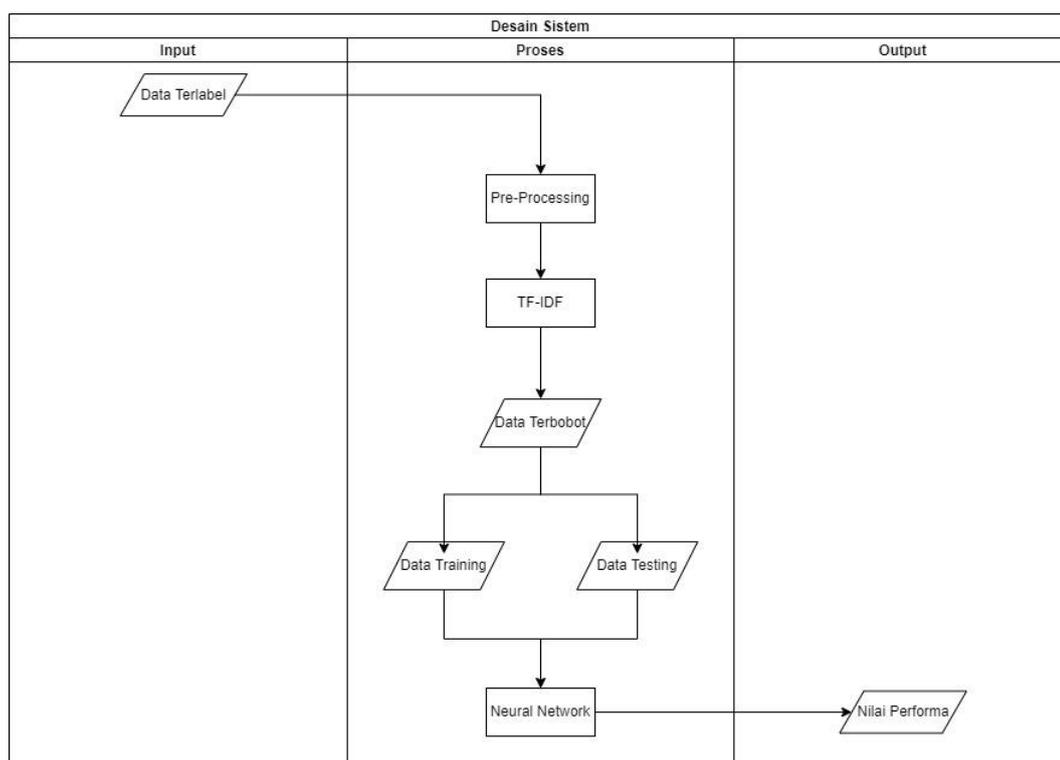
No.	Komentar	Label
1.	semoga aja yang positif tidak melebihi jumlah penduduknya, supaya bisnisnya lancar wkk	Positif

2.	@efren saya setuju pendapat anda dana milyaran wkwkwkwj	Negatif
3.	100000 positif 50000 sembuh setiap naik yg positif yg sembuh tetep 50% tanda tanya bgt...trus yg 50% positif ada di mana? RS kah? Apaiya bisa nampung segitu banyak?kehitung jari RS yg nampung covid ... Logis aja... Dlu ada flu burung indonesi pada rezim itu nolak buat pandemi karena tau itu cuma abal2 dan baik2 saja...kenapa sekarang ikut2an?	Positif
4.	Itahun melawan corona mualai pake masker jaga jarak cuci tangan dll tapi hasil nya preeettt masi banyak yg mati corona masi banyak yg korona masi percaya perinta manusia/ alloh	Positif
5.	Ad yg panen 7 koper uang ,pantasan covid2 tdk mau meredah krn sengaja dibesar2kan agar dana terus mengalir untuk dikorupsi .	Negatif
6.	Ada survei 53% rakyat puas dengan kinerja pemerintah jokowi selama pandemi! Dari segi mananya kita harus berpuas hati? Ditengah jumlah positif yg tiap hari semakin meningkat, ekonomi anjlok dll, bagi mereka yg punya ekonomi diatas fine2 saja, tapi masyarakat yg ekonomi nya rendah sangat2 merasakan dampak corona.	Negatif
7.	Ah aku sdh gk percaya dgn covid. Akal"an aja. Noh ujung"ny ada yg korupsi dr bansos covid mrk dr kubu pdip.	Negatif
8.	Ada uang corona boss makanya virusny ttp dipiara habis uang negara tinggal ngutang g bs bayar ya kasihkan pulau jawa rebes	Negatif
9.	ah beritanya kurang menyenangkan gtu2 j,,yg d tunggu berita kemajuan perekonomian bukan kemajuan copit mlh meledak....	Positif
10.	batuk pilek jare corona. kebanyakan yg kena katanya juga para sultan. yg msarakat mah batuk pilek dari jaman belanda udah ada kale	Negatif

3.3 Desain Sistem

Pada penelitian ini desain sistem menjabarkan alur dari sistem yang akan dibuat. Rancangan dimulai dari pengumpulan data secara *collecting* data kemudian dilakukan pelabelan. Hasil pelabelan akan dilakukan *pre-processing* dengan tahapan yaitu *cleaning*, *case folding*, *tokenizing*, dan *stemming*. Data yang sudah jadi kemudian diberi bobot dan teks diubah menjadi angka menggunakan TF-IDF.

Data berbobot dibagi menjadi dua bagian yaitu data latih dan data uji. Setelah pembagian data selesai, dilanjutkan ke proses klasifikasi komputasi menggunakan metode jaringan syaraf tiruan. Berikut ini adalah diagram perancangan sistem dengan menggunakan flowchart pada Gambar 3.2



Gambar 3. 2 Flowchart Desain Sistem

3.4 Pre-Processing

Tahap ini merupakan sebuah tahapan menyiapkan data yang memiliki peran penting dalam klasifikasi text. Bertujuan untuk menetralsir adanya kesalahan yang dapat mempengaruhi hasil dalam klasifikasi text (Handayani & Pribadi, 2015). Dalam klasifikasi text, *pre-processing* berperan untuk menyaring kata dengan tidak menghilangkan makna kata yang sesungguhnya. Dalam penelitian ini proses *pre-*

processing dilakukan dengan menggunakan Bahasa python. Berikut tahapan dalam *pre-processing* terdiri dari beberapa langkah.

3.4.1 *Cleaning*

Tahap ini dilakukan pembersihan data yang tidak diperlukan dan tidak mempengaruhi kalimat jika dihilangkan. Misalnya dengan menghilangkan tanda baca titik, koma, emoticon, url, hashtag. Dengan melakukan cleansing data akan menjadi lebih bersih dan siap digunakan. Berikut contoh proses cleansing pada Tabel 3.2

Tabel 3. 2 Proses *Cleansing*

Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>
Mau ga mau karena surat vaksin jdi syarat ngurus dokumen penting,,, buah simalakama	Mau ga mau karena surat vaksin jdi syarat ngurus dokumen penting buah simalakama
Pandemi belum usai, yuk terus ikhtiar untuk melawan pandemi ini 😊 Semoga pandemi segera berlalu, agar kehidupan kembali normal lagi	Pandemi belum usai yuk terus ikhtiar untuk melawan pandemi ini Semoga pandemi segera berlalu agar kehidupan kembali normal lagi

3.4.2 *Case Folding*.

Pada proses ini data yang sudah bersih tanpa atribut dan sebagainya masih memiliki kalimat dengan huruf capital. Oleh karena itu dengan *case folding* kalimat tersebut akan diubah menjadi bentuk huruf kecil atau *lower case*. Kalimat yang dapat diubah hanya yang mengandung huruf ‘a’ sampai ‘z’. Berikut contoh proses *case folding* pada tabel 3.3:

Tabel 3. 3 Proses *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Mau ga mau karena surat vaksin jdi syarat ngurus dokumen penting buah simalakama	mau ga mau karena surat vaksin jdi syarat ngurus dokumen penting buah simalakama

Pandemi belum usai, yuk terus ikhtiar untuk melawan pandemi ini Semoga pandemi segera berlalu agar kehidupan kembali normal lagi	pandemi belum usai, yuk terus ikhtiar untuk melawan pandemi ini semoga pandemi segera berlalu agar kehidupan kembali normal lagi
---	---

3.4.3 *Tokenizing*

Pada proses ini adalah proses yang penting dalam teks mining dimana kalimat atau data yang sudah diubah menjadi huruf kecil akan dipecah menjadi perkata. Pemecahannya berdasarkan spasi sehingga diketahui jumlah teks yang mnucul. *Tokenizing* berfungsi agar computer lebih mudah memahami teks yang digunakan. Berikut contoh proses *tokenizing* pada Tabel 3.4

Tabel 3. 4 Proses *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
mau ga mau karena surat vaksin jdi syarat ngurus dokumen penting buah simalakama	“mau”, “ga”, “mau”, “karena”, “surat”, “vaksin”, “jdi”, “syarat”, “ngurus”, “dokumen”, “penting”, “buah”, “simalakama”
pandemi belum usai, yuk terus ikhtiar untuk melawan pandemi ini semoga pandemi segera berlalu agar kehidupan kembali normal lagi	“pandemi”, “belum”, “usai”, “yuk”, “terus”, “ikhtiar”, “untuk”, “melawan”, “pandemi”, “ini”, “semoga”, “pandemi”, “segera”, “berlalu”, “agar”, “kehidupan” “kembali”, “normal”, “lagi”

3.4.4 *Stemming*

Tahap ini dilakukan untuk mengubah kata yang telah dipecah tadi menjadi kata dengan bentuk dasar atau istilah lainnya adalah *root word*. *Stemming* akan menghilangkan kata yang berimbuhan seperti awalan kata, sisipan kata, maupun menghilangkan akhiran kata. Berikut contoh proses *stemming* pada Tabel 3.5

Tabel 3. 5 Proses *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
“mau”, “ga”, “mau”, “karena”, “surat”, “vaksin”, “jadi”, “syarat”, “ngurus”, “dokumen”, “penting”, “buah”, “simalakama”	“mau”, “tidak”, “mau”, “karena”, “surat”, “vaksin”, “jadi”, “syarat”, “urus”, “dokumen”, “penting”, “buah”, “simalakama”
“pandemi”, “belum”, “usai”, “yuk”, “terus”, “ikhtiar”, “untuk”, “melawan”, “pandemi”, “ini”, “semoga”, “pandemi”, “segera”, “berlalu”, “agar”, “kehidupan” “kembali”, “normal”, “lagi”	“pandemi”, “belum”, “usai”, “ayo”, “terus”, “ikhtiar”, “untuk”, “melawan”, “pandemi”, “ini”, “semoga”, “pandemi”, “segera”, “berlalu”, “agar”, “hidup” “kembali”, “normal”, “lagi”

3.5 Word Embedding

Word Embedding merupakan proses yang akan mengonversi teks atau kata menjadi kedalam bentuk vector. Penelitian sebelumnya yang dilakukan oleh (Rahman et al., 2021) menemukan bahwa penyisipan kata lebih efektif dalam mengubah kata menjadi bentuk vektor. Dalam penelitian ini, kami menggunakan metode ini untuk menyematkan kata-kata TF-IDF (*Term Frequency–Inverse Document Frequency*). Proses TF-IDF dapat dilakukan setelah kata atau teks selesai melalui tahap *pre-processing* barulah dapat dilakukan proses TF-IDF dengan melakukan perhitungan nilai frekuensi kemunculan kata. Kemunculan kata ini sangat berpengaruh terhadap besar kecilnya bobot yang akan didapatkan. Semakin banyak kata yang sering muncul, maka nilai bobot yang akan didapatkan juga semakin besar. Berikut rumus perhitungan dan implementasi TF-IDF pada penelitian ini:

$$TF = \frac{\text{jumlah frekuensi kata terpilih}}{\text{jumlah kata}} \quad (3.1)$$

$$IDF = \log e \frac{\text{jumlah dokumen}}{\text{jumlah frekuensi kata terpilih}} \quad (3.2)$$

Tabel 3. 6 Perhitungan TF

Dokumen	Kata	TF
D1	Positif	1
D2		1
D3		0

Tabel 3. 7 Perhitungan IDF

Dokumen	TF	Kata	IDF
D1	1	Positif	$\log \frac{3}{2} = 0.176091$
D2	1		
D3	0		

Tabel 3. 8 Hasil Perhitungan TF-IDF

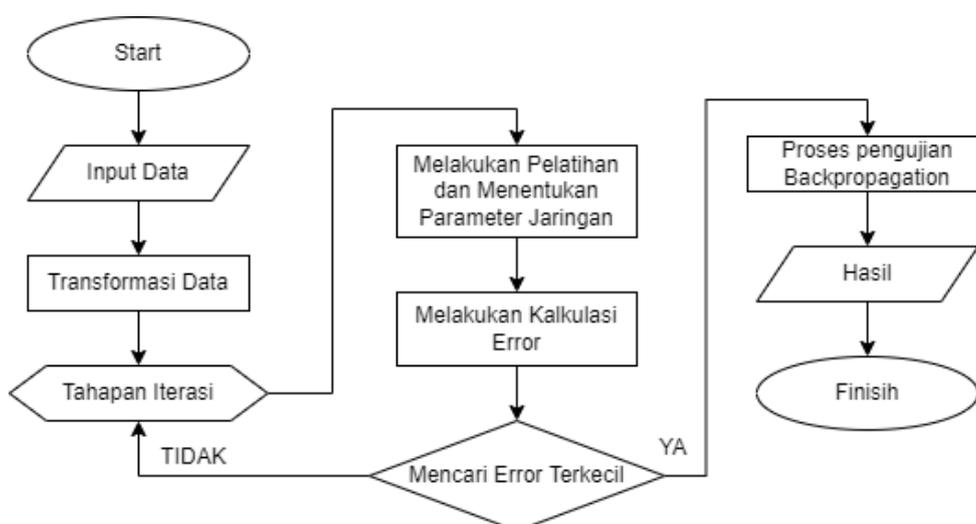
TOKEN	IDF		
	D1	D2	D3
positif	0.176091	0.176091	0
lebih	0.477121	0	0
jumlah	0.176091	0.176091	0
penduduk	0.176091	0.176091	0
supaya	0.477121	0	0
bisnis	0.477121	0	0
lancar	0.477121	0	0
turun	0	0.477121	0
covid	0	0.176091	0.176091
mungkin	0	0	0.477121

belum	0	0	0,477121
hilang	0	0	0,477121
vaksin	0	0	0,477121
semua	0	0	0,477121

Berdasarkan Tabel 3.8 pembobotan dilakukan dengan metode TF-IDF dimana menghasilkan *vector* yang dapat digunakan dalam klasifikasi kalimat dengan metode *neural network*. Dapat dilihat bahwa kata “positif” dan “covid” memiliki nilai terendah yaitu 0,176091 dimana kata tersebut muncul di 2 dokumen dari 3 dokumen yang ada.

3.6 Alur Perhitungan *Neural Network*

Hasil pembobotan yang telah dilakukan dengan TF-IDF selanjutnya akan diproses dengan menggunakan metode *neural network*. Dalam penelitian ini metode *neural network* yang digunakan adalah algoritma *backpropagation*. Sebelum melakukan perhitungan dengan algoritma *backpropagation*, berikut flowchart *backpropagation* :



Gambar 3. 3 Flowchart Algoritma Backpropagation

Berdasarkan Gambar 3.3, dapat dilihat bahwa tahapan dalam menyelesaikan proses dalam algoritma *backpropagation*. Data dalam penelitian ini berasal dari komentar masyarakat di sosial media facebook dalam akun Kemenkes RI pada tahun 2020-2021. Data ini kemudian akan dibagi menjadi dua yaitu data *training* dan data testing. Percobaan perhitungan manual dengan algoritma *backpropagation* disini akan mengambil sampel data berupa kalimat yang telah dilakukan proses pembobotan dengan TF-IDF sebelumnya diatas. Berikut data training yang akan dilakukan proses *backpropagation*:

Kalimat : Dengan vaksin, bismillah tahun depan yang positif covid menurun dan covid segera hilang

Tabel 3. 9 Data Training Hasil Pembobotan

	Vaksin	Positif	Covid	Turun	Hilang
Bobot / kalimat	0.477121	0.176091	0.176091	0.176091	0.477121

Berdasarkan tabel data pembobotan diatas, data tersebut diproses dengan mentransformasikan data ke dalam range [0, 1]. Untuk mentransformasikan data digunakan rumus sebagai berikut :

$$X' = \frac{0.8(x - \min)}{\max - \min} + 0.1 \quad (3.3)$$

Keterangan :

X' : Transformasi Data

x : data yang di transformasikan

min : data paling kecil.

max : data paling besar.

Sehingga dari proses perhitungan transformasi data diatas mendapatkan hasil sebagai berikut :

Tabel 3. 10 Hasil Transformasi Data

x1	x2	x3	x4	x5
0.900001	0.100001	0.100001	0.100001	0.900001

Selanjutnya adalah menentukan learning rate (α), nilai target dan epoch. Nilai epoch diambil secara acak sebagai maksimal iterasi yang akan dilakukan. Nilai learning rate sebesar 0.1. Kemudian memberikan nilai bobot secara acak dari *input* layer menuju *hidden* layer dan nilai bobot acak dari *hidden* layer menuju *output* layer.

Tabel 3. 11 Bobot Nilai Acak *Input* Layer ke *Hidden* Layer

	Z1	Z2
x1	1.5039	-2.0854
x2	-3.1055	-0.6004
x3	-1.4358	-2.1338
x4	1.8041	6.1439
x5	-1.4044	-1.0737
B	4.4769	3.4241

Tabel 3. 12 Bobot Nilai Acak *Hidden* Layer ke *Output* Layer

Y	
Z1	1.7853
Z2	5.3078
b	-3.5465

Selanjutnya dilakukan tahapan iterasi pertama menggunakan algoritma *backpropagation* dilanjutkan dengan fungsi aktivasi sigmoid dengan rumus

sebagai berikut :

$$Z_{net j} = v_{0j} + \sum_{i=1}^p x_i v_{ji}$$

$$z_j = f(Z_{net j}) = \frac{1}{1 + \exp^{-Z_{net j}}}$$

$Z_{net 1} = 4.292728$ Sigmoid $Z_1 = 0.013483$
 $Z_{net 2} = 0.92188$ Sigmoid $Z_2 = 0.2844575$

Perhitungan dilakukan dengan menggunakan Microsoft Excel. Kemudian menghitung keluaran jaringan unit Y dan dilanjutkan menghitung sigmoidnya dengan rumus sebagai berikut :

$$Y_{net1} = -2.01196 \quad Sigmoid Y_{net1} = 0.882047 \quad Y_{net1} = w_{0k} + \sum_{j=1}^n z_j w_{kj}$$

Menghitung factor yang ada dalam setiap unit *output* berdasarkan nilai *error* dengan rumus sebagai berikut:

$$\delta_i = (t_i - y_i) f'(Y_{net1})$$

$$\begin{aligned} \delta_1 &= (0 - 0.882047) * 0.882047 * (0 - 0.882047) \\ &= 0.686239 \end{aligned}$$

Kemudian dilanjutkan dengan menghitung perubahan bobot menggunakan learning rate yaitu 0.1 dengan rumus sebagai berikut :

$$\begin{aligned} \Delta w_0 &= 0.1 * 0.686239 * 0.013483 = 0.000925 \\ \Delta w_1 &= 0.1 * 0.686239 * 0.284575 = 0.019529 \\ \Delta w_2 &= 0.1 * 0.686239 * 0.882047 = 0.060529 \end{aligned} \quad \Delta w_{jk} = \alpha \delta_i z_j$$

Menghitung penjumlahan error dari *hidden* layer sebagai berikut:

$$\begin{aligned} \delta_{net1} &= 0.686239 * 1.7853 = 1.225142 \\ \delta_{net2} &= 0.686239 * 5.3078 = 3.642418 \end{aligned} \quad \delta_{netl} = \sum_{i=1}^k \delta_i w_{kj}$$

Menghitung factor error yang terdapat di *hidden* layer sebagai berikut:

$$\begin{aligned} \delta_i &= \delta_{netl} f'(z_{netj}) = \delta_{netl} z_j (1 - z_j) \\ \delta_1 &= 1.225142 * 1.7853 * (1 - 1.7853) = -1.71764 \\ \delta_2 &= 3.642418 * 5.3078 * (1 - 5.3078) = -83.2837 \end{aligned}$$

Menghitung suku perubahan bobot yang terdapat pada *hidden* layer untuk memperbaiki V_{ij} sebagai berikut: $\Delta v_{ij} = \alpha \delta_j x_i$

Berdasarkan rumus diatas, V_{ij} didapatkan dengan mengalikan hasil δ_1 dengan x1 sehingga menghasilkan tabel sebagai berikut:

Tabel 3. 13 Hasil Perubahan Bobot pada *Hidden* Layer

V1.1	-0.15459	V2.3	-0.83284
V2.1	-7.49554	V1.4	-0.01718
V1.2	-0.01718	V2.4	-0.83284

V2.2	-0.83284	V1.5	-0.15459
V1.3	-0.01718	V2.5	-7.49554

Tahap terakhir dalam metode backpropagation yaitu menghitung perubahan bobot pada *output* layer sebagai berikut :

$$w_{ji}(\text{baru}) = w_{ji}(\text{lama}) + \Delta w_{ji}$$

$$w_1 = 1.7853 + 0.019529 = 1.804829$$

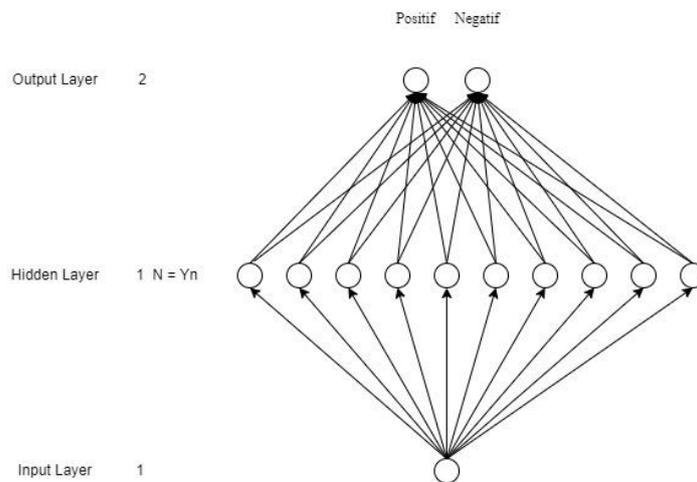
$$w_2 = 5.3078 + 0.060529 = 5.368329$$

$$v_{ji}(\text{baru}) = v_{ji}(\text{lama}) + \Delta v_{ji}$$

Tabel 3. 14 Hasil Vij Baru

V1.1 baru	1.349311912	V3.2 baru	-2.966642355
V1.2 baru	-9.580935231	V4.1 baru	1.78692344
V2.1 baru	-3.12267656	V4.2 baru	5.311057645
V2.2 baru	-1.433242355	V5.1 baru	-1.558988088
V3.1 baru	-1.45297656	V5.2 baru	-1.228288088

Berdasarkan perhitungan manual *neural network* menggunakan algoritma *backpropagation* hasil tabel Tabel 3.14 adalah hasil bobot *hidden* layer dengan iterasi pertama. Hasil tersebut masih belum menunjukkan error minimum yang telah ditetapkan yaitu 0.001. Sehingga dibutuhkan kembali iterasi selanjutnya dengan cara yang sama sampai hasil bobot sesuai dengan *error* yang ditetapkan. Sedangkan perhitungan pada *output* layer mendapatkan nilai sebesar 0.882047 dan nilai error 0.686239 hampir memenuhi nilai target yaitu 1 agar klasifikasi teks dapat digolongkan menjadi kalimat yang positif atau negative. Dapat disimpulkan bahwa dalam perhitungan ini masih memerlukan iterasi selanjutnya untuk mendapatkan hasil yang sebenarnya. Arsitektur jaringan metode *neural network* dalam penelitian ini akan di uraikan pada Gambar 3.4



Gambar 3. 4 Arsitektur Jaringan *Neural Network*

3.7 Skenario Pengujian

Tahapan ini merupakan tahap terakhir dalam penelitian ini yaitu menentukan nilai *accuracy*, *recall*, *precision*, dan *F1-Score*. Sebelum menghitung nilai tersebut, data dibagi menjadi dua yaitu data *training* dan data *testing* masing-masing dengan rasio 90:10. Data *training* merupakan data yang digunakan sebagai data model sedangkan data *testing* digunakan sebagai uji coba model yang akan di stimulasi penggunaan model dunia nyata. Oleh sebab itu data *training* memiliki persentase lebih tinggi daripada data *testing*. Untuk menghitung hasil klasifikasi, penelitian ini menggunakan *confusion matrix* untuk mengenali setiap baris data di kelas yang berbeda. Dalam pengkategorian *confusion matrix* dilakukan klasifikasi data yaitu jika data memiliki nilai benar, maka nilai yang diperoleh adalah *true-positive* dan *true-negative*. Sedangkan data yang memiliki nilai salah memperoleh nilai *false-positive* dan *false-negative*. Berikut tabel 3.15 adalah konsep *confusion matrix*.

Tabel 3. 15 Confusion Matrix

Data		Aktual	
		<i>True</i>	<i>False</i>
Prediksi	<i>False</i>	<i>True Negative (TP)</i>	<i>False Positive (FN)</i>
	<i>True</i>	<i>False Negative (FP)</i>	<i>True Positive (TN)</i>

Berdasarkan Tabel 3.15, berikut pengertian setiap aktual dalam sebuah *confusion matrix*:

1. *True Positive (TP)* adalah data yang jumlahnya memiliki nilai positif dan kemungkinan prediksi positif.
2. *False Positive (FP)* adalah data yang jumlahnya memiliki nilai negative akan tetapi kemungkinan prediksi positif.
3. *False Negative (FN)* adalah data yang jumlahnya memiliki nilai yang positif akan tetapi prediksinya negative.
4. *True Negative (TN)* adalah data yang jumlahnya memiliki nilai yang negative dan prediksi juga negatif.

Berikut konsep rumus menghitung klasifikasi menggunakan *confusion matrix*:

Tabel 3. 16 Rumus Klasifikasi *Confusion Matrix*

Nilai	Rumus
<i>Accuracy</i>	$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$
<i>Precision</i>	$\frac{TP}{TP + FP} \times 100\%$
<i>Recall</i>	$\frac{TP}{TP + FN} \times 100\%$
<i>F1 - Score</i>	$\frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%$

BAB IV

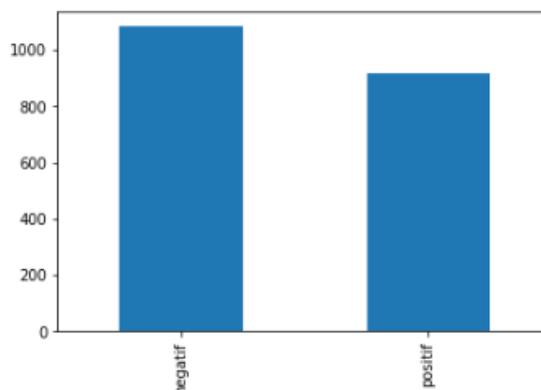
IMPLEMENTASI DAN PEMBAHASAN

Pada bab ini, penulis memaparkan hasil dari uji coba dan pembahasan yang telah dilakukan sesuai dengan tujuan penelitian yang telah di paparkan pada Bab I.

4.1 Skenario Uji Coba

Pembahasan dalam skenario uji coba adalah proses yang diikuti dalam pengembangan sistem penelitian ini. Pengujian yang dilakukan menggunakan notebook Jupyter, Google Colab menggunakan bahasa Python, untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-Score*.

Langkah pertama adalah mendapatkan data berupa komentar pada sosial media dengan *collecting data* yang dilakukan manual *copy-paste* komentar pada akun sosial media Facebook Kepmenkes sebanyak 1999 data. Kemudian data diberikan label berupa positif dan negatif oleh ahli bahasa. Setelah semua data terlabeli proses berikutnya untuk membangun model *Neural Network*. Berikut grafik data yang telah terlabeli berdasarkan kategori positif dan negatif.



Gambar 4. 1 Grafik Data Terlabeli

4.1.1 Tahapan Penelitian

Tahap penelitian ini, dalam membangun model *Neural Network* melalui tahapan-tahapan yang telah di paparkan pada bab sebelumnya.

4.1.1.1 Pre-Processing

Tahap ini terdapat beberapa proses pada data yang telah terlabeli yaitu *cleaning*, *case folding*, *tokenizing*, dan *stemming*. Berikut rangkaian perintah dari *pre-processing*.

a. Cleaning

Pada tahap ini data akan di bersihkan untuk menghilangkan tanda baca, emoticon, url dsb. Dalam proses ini menggunakan library *beautifulsoup* untuk menghapus komentar yang mengandung html, tanda baca dsb.

b. Case folding

Tahap ini data akan di tranformasikan menjadi huruf kecil secara keseluruhan. Tahpa ini sangat penting agar nantinya tidak terjadi kesalahan dalam proses pencarian kata. Fungsi *lower()* digunakan untuk menjadikan seluruh kata menjadi huruf kecil semua.

c. Tokenizing

Pada tahap ini data akan dipecah dari teks menjadi kata atau juga bisa disebut token. Dalam pengimpementasiannya teks dipecah menjadi kata menggunakan fungsi *split()* dalam python.

d. Stemming

Pada tahap stemming ini diperlukan agar tidak banyak variasi kata, dengan mengembalikannya ke kata dasar. Implementasi stemming pada penelitian ini menggunakan library sastrawi. Library sastrawi merupakan library bahasa Indonesia yang digunakan untuk mengubah kata yang berimbuhan menjadi kata dengan bentuk baku atau dasar.

Tahapan *pre-processing* telah dilakukan, berikut contoh sampel data yang telah berhasil dilakukan proses *pre-processing* pada Tabel 4.1

Tabel 4. 1 Sampel Data Hasil Pre-processing

No.	Hasil <i>Pre-processing</i>
1.	[puan, maharani, berikan, bantuan, untuk, anak, yatim, piatu, akibat, pandemi, hati, siapa, tak, pilu, mendengar, kisah, alvi]
2.	[medali, emas, cabang, bulutangkis, yg, diraih, pasangan, ganda, putri, greysia, poliiapriyani, rahayu, pada, olimp]
3.	[ancaman, kelaparan, di, tengah, pandemi, tintasiyasicom, pandemi, covid, di, indonesia, tak]

4.1.1.2 TF-IDF

Pada tahap ini data yang sudah melalui tahap *pre-processing* dilakukan pembobotan dengan TF-IDF. Pembobotan dilakukan dengan menggunakan library TfidfVectorizer.

4.1.1.3 Pembagian Data

Data yang berbobot akan dikelompokkan menjadi dua yaitu data latih dan data uji. Data latih adalah data yang diolah oleh sistem untuk mempelajari sampel data. Sedangkan data uji adalah data yang digunakan untuk menguji model sistem yang berasal dari data latih. Distribusi data

latih dan data uji dilakukan dengan perbandingan 90% untuk data latih dan 10 uji ta. Melakukan dekomposisi data menggunakan train test split dengan input `test_size = 0.1` menunjukkan bahwa data uji adalah 10%. Kemudian menginputkan nilai random state yang merupakan random number generator (RNG) yang boleh diinputkan secara random, disini digunakan inputan nilai 0 agar nilai yang diacak hasilnya sama setiap dijalankan pada program.

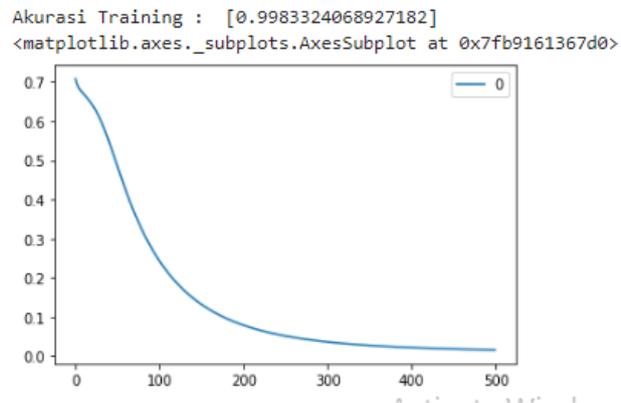
4.1.1.4 Skenario Uji Coba *Neural Network*

Tahap ini, diperlukan inisiasi jumlah node pada *hidden* layer, nilai *epoch*, dan *learning rate* untuk dilakukan uji coba.

- a. Node *hidden layer* : 10-20
- b. Nilai *epoch* : 100-500
- c. Learning rate : 0,0002-0,4

Percobaan yang dilakukan dalam penelitian ini menggunakan inisiasi diatas untuk mendapatkan arsitektur jaringan terbaik untuk klasifikasi kalimat perbincangan masyarakat. Selanjutnya pengimplementasian *neural network* menggunakan fungsi library `MLPClassifier` yang merupakan implementasi jaringan saraf tiruan dengan multi layer perception. Untuk menggunakannya dilakukan install model dengan `from sklearn.neural_network import MLPClassifier`. Kemudian menginputkan inisiasi diatas dengan fungsi aktivasi type logistic yang sama dengan sigmoid untuk klasifikasi dengan dua macam output yaitu 1 dan 0.

Berdasarkan inisiasi yang telah dalam penelitian ini mendapatkan hasil akurasi training berikut.



Gambar 4. 2 Hasil Akurasi Training

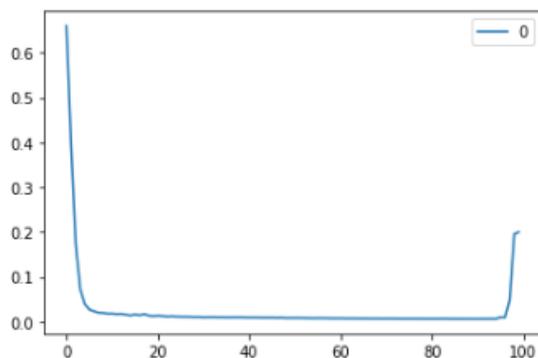
Berdasarkan hasil training pada gambar 4.2, skenario uji coba dilakukan dalam beberapa model untuk mendapatkan model terbaik. Dalam penelitian ini uji coba dengan 5 model, dimana diberikan label untuk membedakan model yang akan dilakukan dengan label model A-E. Uji coba model tersebut dibedakan atributnya dari segi hidden layer, neuron, learning rate, dan epoch yang bervariasi untuk menghasilkan jaringan dengan model terbaik. Berikut uji coba model *neural network* dalam penelitian ini pada tabel 4.2.

Tabel 4. 2 Daftar Pengujian Model Neural Network

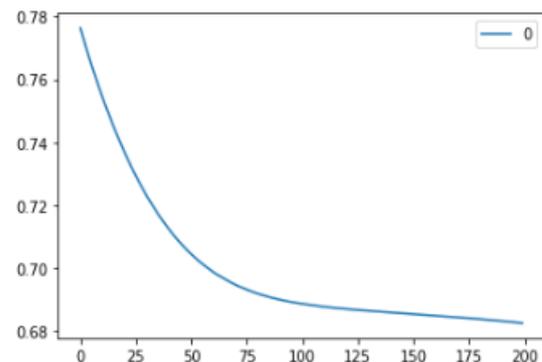
Model	Hidden Layer	Neuron	Learning Rate	Epoch	Akurasi Training
Model A	1	10	0,1	100	0.9794
Model B	3	6,14,5	0,0002	200	0.5541
Model C	2	20,5	0,003	300	0.9983
Model D	2	8,3	0,4	400	0.9922
Model E	1	19	0,01	500	0.9983

Model A memiliki hidden layer 1, neuron 10, learning rate 0.1, dan nilai epoch 100 yang menghasilkan akurasi training sebesar 0.9794.

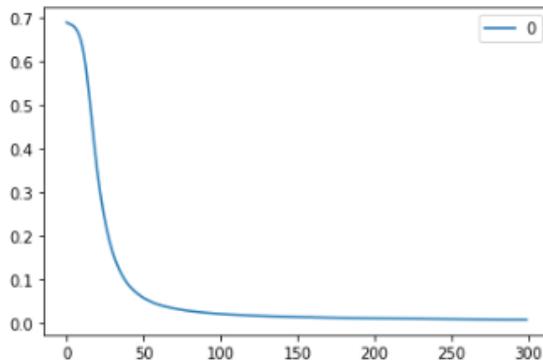
Model B memiliki hidden layer 3(neuron 6,14,5), learning rate 0.0002 dan nilai epoch 200 yang menghasilkan akurasi training sebesar 0.5541. Model C memiliki hidden layer 2 (neuron 20,5), learning rate 0.003, dan nilai epoch 300 yang menghasilkan akurasi training sebesar 0.9983. Model D memiliki hidden layer 2 (neuron 8,3), learning rate 400, dan nilai epoch 400 yang menghasilkan akurasi training 0.9922. Model E memiliki hidden layer 1, neuron 19, learning rate 0.01, dan nilai epoch 500. Berdasarkan uji coba model pada Tabel 4.2, model E merupakan model yang paling tinggi nilai akurasi trainingnya dan memiliki grafik yang lebih stabil dibandingkan dengan model C. Hal ini dapat dibuktikan dengan grafik pada gambar 4.4 hingga gambar 4.7 untuk membandingkan dan memilih model paling konvergen dan cocok di implementasikan. Berikut grafik *loss curve* dan *epoch* pada setiap model.



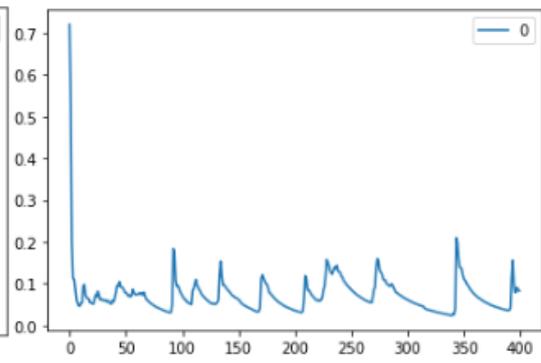
Gambar 4. 4 Grafik Pada Model A



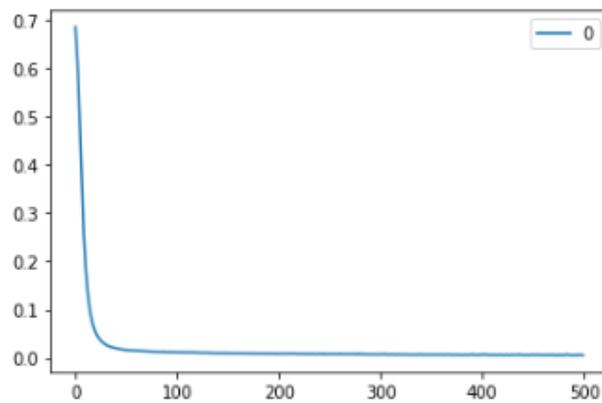
Gambar 4. 3 Grafik Pada Model B



Gambar 4. 6 Grafik Pada Model C



Gambar 4. 5 Grafik Pada Model D

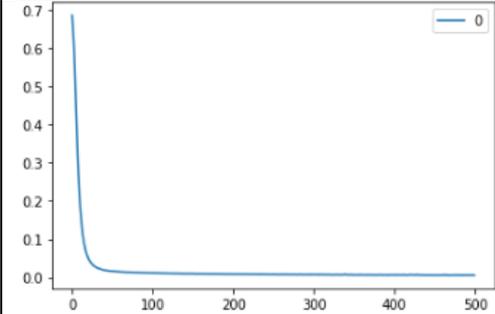
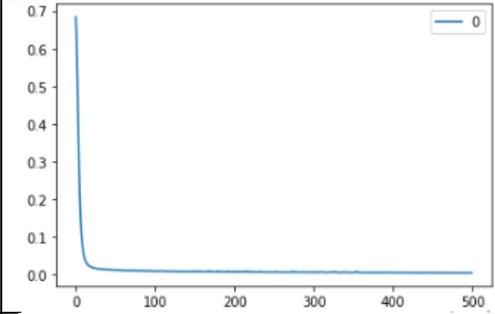
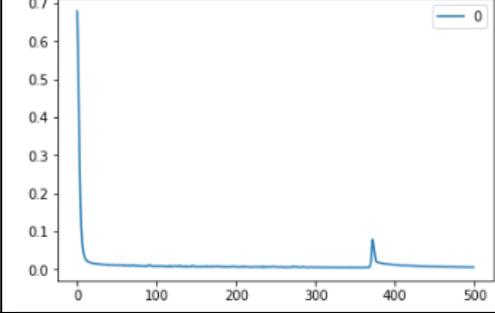
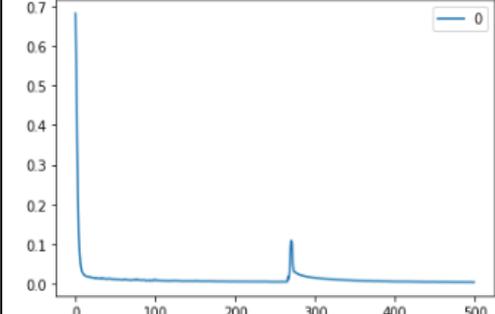


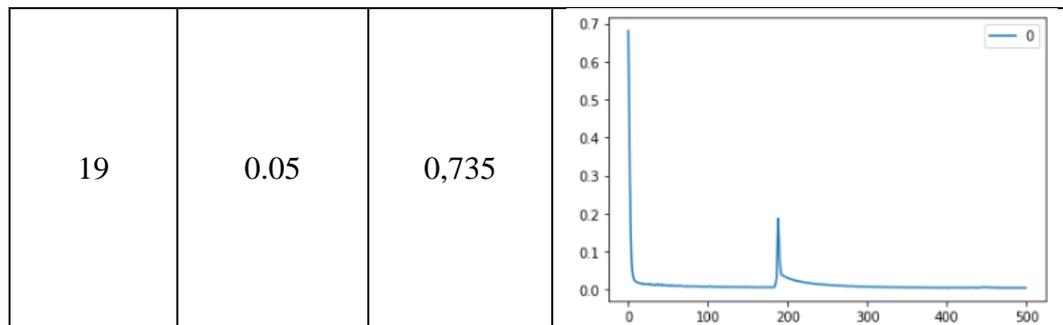
Gambar 4. 7 Grafik Pada Model E

Beberapa percobaan model yang telah dilakukan, grafik pada gambar 4.4 sampai gambar 4.7 menyatakan dengan menggunakan 1 *hidden layer* dengan neuron 19 mendapatkan nilai loss terendah serta stabil dan akurasi training tertinggi pada model E. Sehingga penelitian ini implementasi model yang digunakan untuk metode neural network adalah model E. Berdasarkan skenario uji coba model yang telah dilakukan model E adalah model yang terbaik. Sehingga dalam penelitian ini menggunakan model E yang memiliki 1 *hidden layer* dengan node yang berjumlah 19 node dimana parameter learning ratenya akan diuji kembali

dengan learning rate 0.01-0.05 pada table 4.3 untuk mencari learning rate yang sesuai dan mendapatkan nilai akurasi testing tertinggi.

Tabel 4. 3 Skenario Uji Coba Model E

Jumlah Node	Learning Rate	Nilai Akurasi	Grafik
19	0.01	0,755	
19	0.02	0,745	
19	0.03	0,735	
19	0.04	0,72	



4.2 Hasil Uji Coba

Hasil pengujian dari penelitian ini bertujuan untuk mendapatkan nilai performansi terbaik dari sistem yang dihasilkan. Dalam penelitian ini data yang digunakan adalah data tahun 1999 dimana data tersebut akan dibagi menjadi dua yaitu data latih dan data uji dengan perbandingan 90:10. Data latih yang diuji adalah 1799 data, sedangkan data uji yang diuji adalah 199 data. Selanjutnya dilakukan pengukuran nilai performa berupa *accuracy*, *precision*, *recall*, dan *F1-Score*. Berdasarkan skenario uji coba model yang telah dilakukan model E adalah model yang terbaik. Sehingga dalam penelitian ini menggunakan model E yang memiliki 1 *hidden layer* dengan node yang berjumlah 19 node dimana parameter learning ratenya akan diuji kembali dengan learning rate 0.01-0.05. Uji coba model E tersebut menghasilkan nilai akurasi testing tertinggi dengan menggunakan learning rate 0.01.

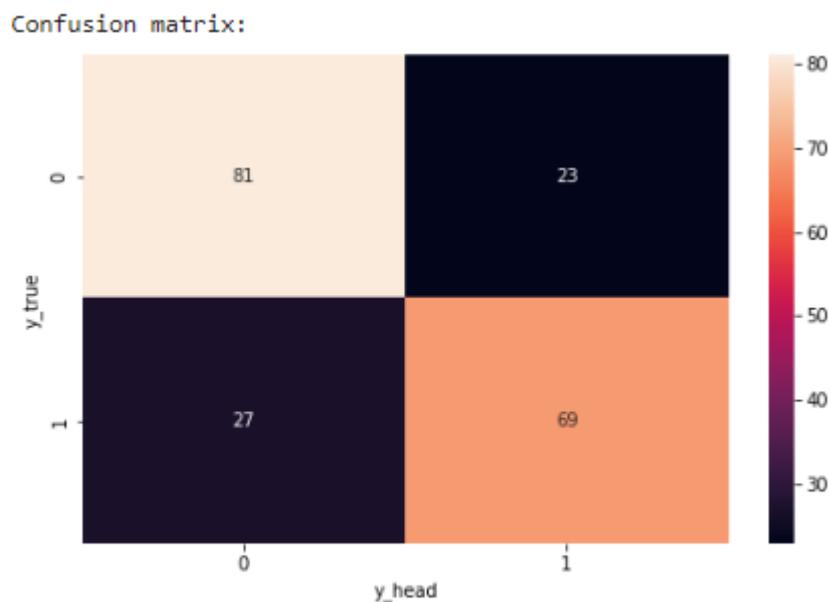
Nilai akurasi testing tertinggi adalah 0,755 dan nilai akurasi training sebesar 0.9983. Hasil tersebut diperoleh dengan menghitung nilai y_{train} dan y_{pred} dalam klasifikasi menggunakan metode *neural network*. Selanjutnya dilakukan perhitungan nilai performa pada data testing menggunakan *confusion matrix*. Untuk mendapatkan nilai matrix digunakan model sklearn atau scikit-learn untuk

membantu pemrosesan data classification report dan *confusion matrix* digunakan untuk mendapatkan nilai pengukuran system. Berikut hasil matrix dan nilai performa *accuracy*, *precision*, *recall*, dan *F1-Score*.

```
[[81 23]
 [27 69]]
```

	precision	recall	f1-score	support
negatif	0.75	0.78	0.76	104
positif	0.75	0.72	0.73	96
accuracy			0.75	200
macro avg	0.75	0.75	0.75	200
weighted avg	0.75	0.75	0.75	200

Gambar 4. 8 Nilai Akurasi Testing



Gambar 4. 9 Hasil Confusion Matrix

Gambar 4.8 merupakan hasil pengukuran nilai performa pada sistem menggunakan confusion matrix. Gambar 4.9 merupakan nilai matrix yang akan digunakan sebagai acuan dalam menghitung nilai performa. 81 merupakan nilai

True Positive (TP). 23 merupakan nilai False Negative (FN). 27 merupakan nilai False Positive (FP). 69 merupakan nilai True Negatif (TN). Hasil pengukurannya didapatkan *accuracy* sebesar 75%, *precision* 75%, *recall* 78%, dan F1-Score 76%. Berikut merupakan perhitungan manual sebagai pembuktian untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{81 + 69}{81 + 69 + 27 + 23} \times 100\% \\
 &= \frac{150}{200} \times 100\% = 75\%
 \end{aligned} \tag{4.1}$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{81}{81 + 27} \times 100\% \\
 &= \frac{80}{108} \times 100\% = 75\%
 \end{aligned} \tag{4.2}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{81}{81 + 23} \times 100\% \\
 &= \frac{80}{104} \times 100\% = 78\%
 \end{aligned} \tag{4.3}$$

$$\begin{aligned}
 \text{F1-Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \\
 &= \frac{2 \times 0.75 \times 0.78}{0.75 + 0.78} \times 100\% = 76\%
 \end{aligned} \tag{4.4}$$

Berdasarkan perhitungan manual tersebut menunjukkan bahwa klasifikasi menggunakan *neural network* dengan skenario uji coba *confusion matrix* adalah baik dalam memprediksi klasifikasi baik positif dan negatif. Selanjutnya

pengimplementasian klasifikasi kalimat, Seperti yang telah dijelaskan pada bab sebelumnya dari penelitian ini, data akan dibagi menjadi dua yaitu data latih dan data uji. Data latih merupakan data yang akan diolah untuk mendapatkan model jaringan syaraf tiruan yang akan disimpan menggunakan active pickle library untuk menyimpan model data tersebut. Model yang tersimpan akan digunakan untuk proses klasifikasi kalimat perbincangan masyarakat. Kemudian model yang telah disimpan digunakan untuk klasifikasi kalimat dengan mencoba data dari collecting data di sosial media yang telah dilabeli sehingga menghasilkan klasifikasi kalimat termasuk negatif atau positif yang sama dengan data yang telah dilabeli.

```
new_data = input("Input Data Baru: ")
predict_new_text(new_data)

Input Data Baru: Secepatnya pandemi covid berakhir di Indonesia.. Aamiin..
'Hasil Klasifikasi Positif 95.83%'
```

Gambar 4. 10 Hasil Klasifikasi Kalimat Perbincangan di Sosial Media

Pada Gambar 4.10, percobaan klasifikasi kalimat dengan menginputkan data atau komentar baru sehingga dapat diketahui bahwa komentar tersebut tergolong dalam klasifikasi kalimat positif. Berikut tabel 4.5 merupakan data hasil label asli dari ahli bahasa dan hasil prediksi dari sistem, dimana data yang di cantumkan merupakan sampel data testing sebanyak 40 data untuk menguji system.

Tabel 4. 4 Hasil Klasifikasi Sampel Data

No	Data	Label		
		Input	Klasifikasi	Persentase
1	Alhamdulillah yg sembuh semakin meningkat ...	Positif	Positif	100.00%

2	Secepatnya pandemi covid berakhir di Indonesia.. Aamiin..	Positif	Positif	95.83%
3	Alhamdulillah ya Allah yg sembuh semakin banyak dibanding yg positif	Positif	Positif	100.00%
4	Semangat buat tenaga medis yg selalu menjadi garda terdepan dalam penanganan covid-19 di indonesia	Positif	Positif	99.92%
5	Terima kasih yang telah mematuhi protokol kesehatan Lekas sembuh dan segera pergi dari bumi ku	Positif	Positif	100.00%
6	Alhamdulillah, hari ini banyakan yang sembuh.	Positif	Positif	100.00%
7	Klo yg sembuh setiap hari selalu lebih banyak dari yg terinfeksi kayanya virus di indonesia semakin menipis	Positif	Positif	99.44%
8	Alhamdulillah yg meninggal dunia semakin berkurang	Positif	Positif	99.74%
9	Alhamdulillah peningkatan kesembuhan makin banyak	Positif	Positif	100.00%
10	Pembohongan publik, udah pake masker udah jaga jarak , tapi tetep aja nambah, ini konspirasi	Negatif	Negatif	100.00%
11	Obat belum ketemu yg sembuh bnyak.. aneh.	Negatif	Negatif	77.20%
12	Alhamdulillah di kota Serang kuat iman tidak terpengaruh oleh virus Corona.	Positif	Positif	99.95%

13	Habis deh orang Indonesia..... lama kelamaan	Negatif	Negatif	99.98%
14	Selalu ikuti anjuran pemerintah n protokol kesehatan n jgn lupa berdo'a...	Positif	Positif	99.99%
15	Akhir zaman penuh fitnah	Negatif	Negatif	100.00%
16	Alhamdulillah yang sembung lebih banyak daripada yang positif	Positif	Positif	100.00%
17	Alhamdulillah covid 19 akan segera berakhir	Positif	Positif	97.40%
18	Ya Allah lindungilah rakyat Indonesia dari covid 19	Positif	Positif	99.88%
19	Yeayy, banyak orang meninggal karena korona, malaikat panen	Negatif	Negatif	99.32%
20	Gara-gara covid banyak yang meninggal kan rasainn	Negatif	Positif	90.54%
21.	Alat test rapid nya di bakar pasti korona hangus	Negatif	Negatif	99.59%
22.	Berhubung saya ga punya stiker nya jadi di ketik aja "kami se-kabupaten mengucapkan BODO AMAT"	Negatif	Negatif	100.00%
24.	Betul boss tempat ku jg sma anak2 muda msih nongkrong sperti biasa	Negatif	Negatif	99.82%
25.	Hanya ada satu penyakit di INDONESIA sekarang yaitu corona.	Negatif	Negatif	99.99%
26.	Indonesia bakalan miskin kalau terus menerus ikut dalam permainan corona	Negatif	Negatif	100.00%
27.	tiap hari makin nambah aja yg positif kapan wabah ini mereda	Negatif	Negatif	100.00%

28.	Mati kerana sakit Jatung. Di Bilang mati covid19... Tuhan sudah mengatur kapan kita akan pergi!	Negatif	Negatif	99.95%
29.	Astaghfirullah semoga cepat sembuh Indonesia ku dan masyaraka Nya bisa beraktivitas seperti biasanya tapi saya agak kecewa pada Indonesia kenapa susu sapi iklannya naga tapi namanya beruang padahal kan itu susu sapi bukan naga ataupun beruang	Positif	Positif	100.00%
30.	Ini dokter atau org yg buat ttg aturan jaga jarak.. Jaga jarak dari org sakit bukan dari org yg sehat.. Jd yg merasa sakit sebaik nya sadar dan jaga jarak	Positif	Positif	99.37%
31.	Saran kpd petugas vaksin dan pemerintah daerah utk kelompok lansia spy cakupannya tinggi hrs seperti jemput bolah, kl perlu sampai dor to dor.	Positif	Positif	99.56%
32.	Semoga pandemi covid19 cepat hilang dimuka bumi ttp jaga kesehatan ya semua ttp pakai masker taati prokes	Positif	Positif	100.00%
33.	Tetap patuhi protokol kesehatan dan PPKM agar pandemi cepat berakhir terima kasih atas pengorbanan para medis yang berjuang tanpa henti	Positif	Positif	100.00%

34.	Alhamdulillah angka positif dan meninggal sdh menurun sedang angka sembuh meningkat Semoga grafik yg bagus berlanjut di hari-hari berikutnya Aamiin..YRA	Positif	Positif	100.00%
35.	yo jalankan protokol kesehatan secara benar dan disiplin demi melindungi diri dan orang-orang yang kita sayangi	Positif	Positif	100.00%
36.	Benar , pandemi bisa berkurang tergantung diri sendiri mentaati protokol kesehatan dan vaksin	Positif	Positif	100.00%
37.	Tunggu aja pendeminya pas ada pemilu ...pasti akan berakhirtapi jika yg menang dari kaum mereka ya g tau lagi.	Negatif	Negatif	100.00%
38.	Vaksin itu nggak berguna bagi orang petani yang berguna paling celurit untuk ngambil daun-daunan buat makan kambing kambing saya sakit coba divaksin	Negatif	Negatif	99.67%
39.	ya jelas lh Indonesia jdi nomor 1, yg gk covid aja di covidkan, dri situ aja dh brp ribu org, tambah lgi yg mati	Negatif	Negatif	99.99%
40.	yg mengkofivkan Pasion di rumah itu .bukan penyakit Corona ..tapi si alat rapat tes..krna alat rapat tes itu adalah	Negatif	Negatif	99.99%

	dinamik..hasilnya pasti berubah			
--	---------------------------------	--	--	--

4.3 Pembahasan

Berdasarkan hasil uji coba yang telah dilakukan dalam penelitian ini menggunakan metode *neural network* pada kalimat perbincangan masyarakat di sosial media fecaebook tentang Covid-19, pembagian data dengan perbandingan 90:10 yaitu 1799 data training dan 199 data testing. Data tersebut diklasifikasikan menjadi dua kategori yaitu kalimat positif dan negatif yang telah dilabeli oleh ahli bahasa dengan memberikan sampel sebanyak 10 data kemudian, sampel tersebut dibuat formula yang digunakan untuk pelabelan selanjutnya oleh peneliti. Selanjutnya data akan melalui *pre-processing* dan TF-IDF untuk pembersihan data dan pembobotan. Proses selanjutnya adalah pemodelan menggunakan metode *neural network* dimana dilakukan beberapa pemodelan untuk mendapatkan model terbaik. Dalam proses training menggunakan *neural network* di pengaruhi oleh jumlah *epoch* yang di tentukan. Semakin tinggi nilai epoch akan mempengaruhi hasil akhirnya untuk menghasilkan *loss* terkecil. Akan tetapi apabila nilai *epoch* terlalu sedikit memiliki dampak pada nilai bobot di setiap nodenya tidak akan di pengaruhi dengan maksimal sehingga nilai *error* yang dihasilkan semakin besar.

Sebaliknya, jika jumlah epoch terlalu besar akan menyebabkan *overfitting* yang berasal dari proses training sehingga akan membutuhkan waktu sedikit lebih lama untuk menghasilkan modelnya. Selain itu pemodelan melalui uji coba yang beragam juga berpengaruh pada sistem. Oleh karna itu scenario uji coba yang beragam juga penting untuk mendapatkan hasil model yang optimal dan konvergen. Berdasarkan scenario uji coba yang dilakukan dengan menggunakan data

sebanyak 1999 data berupa kalimat perbincangan masyarakat dengan inisiasi yang bervariasi sehingga mendapatkan model E menggunakan learning rate 0.01 menunjukkan bahwa klasifikasi yang dilakukan dapat berjalan dengan baik. Hasil pengukuran nilai performa menggunakan *confusion matrix* mendapatkan nilai 81 True Positive (TP). 23 False Negative (FN). 27 False Positive (FP). 69 True Negatif (TN). Nilai tersebut digunakan untuk menghitung nilai performa sistem menggunakan rumus yang telah dijelaskan pada bab 3. Sehingga dari perhitungan tersebut mendapatkan nilai *accuracy* sebesar 75%, *precision* 75%, *recall* 78%, dan *F1-Score* sebesar 76%. Hasil klasifikasi kalimat perbincangan masyarakat ini dapat digunakan oleh pemerintah sebagai evaluasi dalam menetapkan kebijakan untuk menangani Covid-19 hal ini disebabkan data yang dikumpulkan terdapat kalimat negatif yang ditunjukkan kepada pemerintah dan kalimat positif sebagai penyemangat kepada masyarakat dan tenaga kesehatan yang sedang mengalami dan menangani Covid-19.

Masyarakat yang memberikan pendapat berupa komentar di sosial media sering kali menuaikan pendapatnya dengan berkata-kata yang kurang baik sampai menyudutkan bahkan tidak mempercayai adanya Covid-19. Sebagai warga negara Indonesia yang mayoritas penduduknya Islam seharusnya mereka dapat menyampaikan pendapatnya dengan bijak dan tutur kata yang baik. Sebagai seorang muslim yang baik kita sudah diberikan aturan oleh Allah *سُبْحَانَہٗ وَ تَعَالَى* dalam menyampaikan suatu perkataan seperti dalam firman-Nya dalam Qs. Al- Ahzab : 70-71 yang berbunyi :

يَأَيُّهَا الَّذِينَ ءَامَنُوا اتَّقُوا اللَّهَ وَقُولُوا قَوْلًا سَدِيدًا ﴿٧٠﴾ يُصْلِحْ لَكُمْ
 أَعْمَالَكُمْ وَيَغْفِرْ لَكُمْ ذُنُوبَكُمْ وَمَنْ يُطِيعِ اللَّهَ وَرَسُولَهُ فَقَدْ فَازَ فَوْزًا
 عَظِيمًا ﴿٧١﴾

“Hai orang-orang yang beriman, bertakwalah kamu kepada Allah dan katakanlah perkataan yang benar. niscaya Allah memperbaiki bagimu amalan-amalanmu dan mengampuni bagimu dosa-dosamu. Dan barangsiapa mentaati Allah dan Rasul-Nya, maka sesungguhnya ia telah mendapat kemenangan yang besar”. (QS. Al-Ahzab : 70-71)

Dalam ayat 70 Allah memerintahkan hambanya yang beriman untuk senantiasa bertakwa dan menyembah Nya. Ayat tersebut juga memerintahkan kepada hamba yang beriman untuk selalu menyampaikan perkataan yang benar, jujur, tidak berbelit-belit, dan tidak pula menyimpang. Dalam ayat 71 Allah berjanji akan memberi pahala dengan memperbaiki amalan dan mengampuni dosa-dosa orang yang beriman, bertakwa, dan menyampaikan perkataan yang benar. Memperbaiki amalan yang dimaksudkan adalah Allah akan memberi taufik kepada mereka untuk mengerjakan amalan-amalan yang baik dan mengampuni dosa-dosa mereka terdahulu. Selain itu, Allah juga akan memberi mereka ilham untuk bertaubat agar terhindar dari dosa di kemudian hari.

Imam Baidhawi dalam tafsirnya menjelaskan bahwa makna “qoulang syadiidaa” pada ayat 70 adalah untuk melindungi perkataan yang akan mendatangkan manfaat tertentu. Pertama, membimbing ruh untuk berbuat baik agar hidupnya selalu dalam lindungan Tuhan dan menjadikan amalnya diterima oleh

Tuhan. Kedua, Tuhan akan mengampuni dosa-dosanya karena Dia selalu berada di sisinya dalam perkataan dan perbuatan.

Dari surat Al-Ahzab ayat 70-71 kita mengetahui pentingnya menjaga tutur kata baik secara lisan dan juga secara tertulis pada platform sosial media. Sehubungan dengan perintah tersebut, kita sebagai umat muslim berkewajiban bertutur kata yang baik dalam menyampaikan sesuatu di sosial media. Sehingga ketika kita kurang setuju terhadap kebijakan pemerintah dalam menangani Covid-19, tidak langsung berkomentar negatif dengan memikirkan terlebih dahulu keefektifan kebijakan tersebut bagi masyarakat. Selain itu, sebagai umat muslim dalam bertutur kata yang baik juga berkewajiban menyampaikan perkataan yang benar dan sesuai fakta yang ada tidak melebih-lebihkan. Hal ini disebabkan nantinya setiap perbuatan yang dilakukan oleh manusia akan di mintai pertanggung jawaban atas apa yang telah dilakukannya. Seperti Firman Allah dalam QS. Al-Zalzalah : 7-8.

فَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ خَيْرًا يَرَهُ ﴿٧﴾ وَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ شَرًّا يَرَهُ ﴿٨﴾

“Barangsiapa yang mengerjakan kebaikan seberat dzarrahpun, niscaya dia akan melihat (balasan)nya. Dan barangsiapa yang mengerjakan kejahatan sebesar dzarrahpun, niscaya dia akan melihat (balasan)nya pula.” (QS. Al-Zalzalah : 7-8)

Dalam ayat tersebut, Allah menyebutkan semua amal pasti akan mendapat balasan masing-masing. Ayat 7 menjelaskan amal baik, walaupun hanya seberat dzarrah niscaya akan diberikan balasan. Ayat 8, amal jahat walaupun hanya seberat

dzarrah akan diberikan balasan. Menurut tafsir Al-Wajiz (TafsirWeb, n.d.), barang siapa yang melakukan keburukan pada saat di dunia meskipun seberat biji dzarrah, maka nanti akan diketahui balasannya diakhirat. Dalam perkataannya sendiri, Ibnu Abu Hatim dari Sa'id bin Jubair mengatakan bahwa “Ketika ayat ini diturunkan, setiap muslim mengira bahwa dia tidak akan mendapatkan pahala yang kecil atas sesuatu jika diberikan kepada orang lain, sedangkan ada juga yang berpikir bahwa mereka akan melakukan hal-hal sepele seperti berbohong, bersumpah, mengintip, dll. Sungguh, Allah tidak menjanjikan neraka hanya untuk mereka yang melakukan kejahatan berat, itulah sebabnya dua hukuman ini diturunkan. Terdapat juga istilah Amar makruf nahi munkar merupakan istilah fiqh dengan sebutan al-Hisbah. Istilah tersebut merupakan perintah yang ditujukan kepada seluruh masyarakat untuk selalu mengajak dalam hal berperilaku yang baik dan mencegah untuk tidak melakukan perilaku yang buruk. Amar makruf nahi munkar hukumnya wajib bagi umat Islam, dimana jika tidak menjalaninya akan mendapatkan dosa dan hukuman berupa siksaan di akhirat nanti. Dalam al-quran mar makruf nahi munkar juga disinggung dalam QS. Al-Imran :110 yang memiliki arti berikut:

“Kamu adalah umat yang terbaik yang dilahirkan untuk manusia, menyuruh kepada yang ma'ruf, dan mencegah dari yang munkar, dan beriman kepada Allah....”

Ayat tersebut dalam bukunya oleh Muhammad Abdul Qadir mengatakan bahwa amar makruf adalah sebuah perintah yang dijalankan untuk hal kebaikan, sedangkan munkar merupakan dimana setiap kebaikan, perkataan, dan perbuatan yang diingkari oleh Allah harus di jauhi. Jika dikaitkan dengan klasifikasi

perbincangan kalimat di sosial media ini kita sebagai umat muslim dalam berkomentar dengan bertutur kata yang baik dan menyampaikan dengan fakta yang ada.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil uji coba yang telah dilakukan hasil klasifikasi kalimat perbincangan masyarakat menggunakan *Neural Network* yang dilakukan dengan melalui beberapa tahapan diantaranya pengumpulan data berupa teks yang diambil secara collecting data dari sosial media facebook berupa komentar di akun Kepmenkes RI. Selanjutnya dilakukan proses pelabelan oleh ahli bahasa dengan kategori kalimat positif dan negatif. Selanjutnya masuk kepada proses pengolahan kata dengan *pre-processing* agar data mentah yang telah diambil telah siap untuk digunakan dalam penelitian. Tahap *pre-processing* memiliki 4 tahapan yaitu cleaning(membersihkan data), case folding(merubah kalimat menjadi huruf kecil), tokenizing(mengubah kalimat menjadi kata), stemming(mengubah kata menjadi kata dasar). Setelah proses dilakukan, selanjutnya data akan di transformasikan dengan pembobotan pada masing-masing kata menggunakan TF-IDF. Setelah pembobotan dilakukan data akan dibagi menjadi dua yaitu data training dan data testing dengan perbandingan 90:10. Proses selanjutnya pada penelitian ini adalah klasifikasi kalimat perbincangan masyarakat menggunakan metode *neural network*.

Pengimplementasian *Neural Network* menggunakan fungsi library `MLPClassifier` yang merupakan implementasi jaringan saraf tiruan dengan multi layer perception. Untuk menggunakannya dilakukan install model dengan `from sklearn.neural_network import MLPClassifier`. Kemudian menginputkan

inisiasi diatas dengan fungsi aktivasi type logistic yang sama dengan sigmoid untuk klasifikasi dengan dua macam output yaitu 1 dan 0. Implementasi metode *neural network* dilakukan dengan beberap skenario uji coba untuk mendapatkan model jaringan terbaik dengan atribut berbeda-beda sehingga dihasilkan model E yang memiliki hidden layer 1 , neuron 19, learning rate 0.01 dan nilai epoch 500. Model E memiliki nilai akurasi testing tertinggi dan grafk yang optimal sehingga model E inilah yang kemudian dilakukan pengujian terhadap untuk menghitung nilai performa berupa *accuracy*, *precision*, *recall*, dan *F1-Score* dengan meggunakan *Confusion Matrix*. Berdasarkan hasil pengujian terhadap klasifikasi kalimat perbincangan masyarakat mendapatkan hasil *accuracy* 75%, *precision* 75%, *recall* 78%, *F1-Score* 76%. Hal ini menunjukkan bahwa banyaknya node pada *hidden layer*, learning rate, epoch mempengaruhi hasil klasifikasi kalimat perbincangan masyarakat menggunakan metode *neural network*.

5.2 Saran

Dalam penelitian ini peneliti menyadari bahwa hasil dari penelitian ini masih belum sempurna. Oleh karena itu peneliti menyarankan melakukan beberapa hal untuk meningkakan hasil dari performa (*accuracy*, *precicy*, *recal*, *F1-Score*) berikut:

1. Menambah data yang digunakan serta atribut untuk menghasilkan nilai performa akurasi lebih tinggi.
2. Pembobotan selain TF-IDF misalnya Word2Vec dan pembobotan lain.

3. Mencoba untuk melakukan dengan skenario pengujian lainnya untuk mendapatkan nilai performa yang lebih baik
4. Pengklasifikasian dapat menggunakan lebih dari dua macam klasifikasi
5. Penelitian selanjutnya sebaiknya dilakukan dengan menggunakan bahasa selain Bahasa Indonesia atau menggunakan Bahasa Asing.

DAFTAR PUSTAKA

- Aditya Jain, Gandhar Kulkarn, V. S. (2018). Natural Language Processing. *International Journal of Computer Sciences and Engineering*, 6(10), 628–632. <https://doi.org/10.26438/ijcse/v6i1.161167>
- Almuttaqi, A. I. (2020). Kekacauan Respons terhadap COVID-19 di Indonesia. *The Habibie Center Insights*, 1(13), 1–7. <http://habibiecenter.or.id/img/publication/66f28c42de71fefe1c6fcdee37a5c1a6.pdf>
- Ananto, M. I., Winahju, W. S., & Fithriasari, K. (2019). Klasifikasi Kategori Pengaduan Masyarakat Melalui Kanal LAPOR! Menggunakan Artificial Neural Network. *Inferensi*, 2(2), 71. <https://doi.org/10.12962/j27213862.v2i2.6821>
- Bima, R., Ardika, P., Irawan, B., Si, S., Setianingsih, C., Prodi, S., Komputer, T., Elektro, F. T., & Telkom, U. (2020). *Analisis Sentimen Data Pada Bpjs Kesehatan Dengan Metode Backpropagation Neural Network Sentiment Analysis of Bpjs Kesehatan Data With Backpropagation Neural Network Method*. 2–8.
- Choldun, M. I., & Surendro, K. (2018). Klasifikasi Penelitian Dalam Deep Learning. *Improve*, 10(1), 25–33. <https://ejurnal.poltekpos.ac.id/index.php/improve/article/view/338>
- Cholissodin, I., & Soebroto, A. A. (2020). *AI, MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi)*. July 2019.
- Deng, L., & Yu, D. (2013). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4), 197–387. <https://doi.org/10.1561/20000000039>
- Euis Saraswati, Yuyun Umidah, A. V. (2021). Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19. 9(1), 56–60.
- Fausett, L. (2004). *Fundamentals of Neural Network : Architectures, Algorithms, and Applications*. Prantice Hall.
- Firdaus, A., Firdaus, W. I., Studi, P., Informatika, T., Digital, M., & Sriwijaya, P. N. (2021). *Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan) Ali*. 13(1), 66–78.
- Grosz, B. J. (1982). Natural language processing. *Artificial Intelligence*, 19(2), 131–136. [https://doi.org/10.1016/0004-3702\(82\)90032-7](https://doi.org/10.1016/0004-3702(82)90032-7)
- Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–

24. <https://doi.org/10.15294/jte.v7i1.8585>

- Hestina, N. A. (2020). Wabah Penyakit Menular (COVID-19) dan Perumpamaan Dalam Al-Qur'an. *Studi Al-Qur'an Dan Keislaman*, 4(02), 125–13.
- Hidayat, E. Y., Hardiansyah, R. W., & Affandy, A. (2021). Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 108–118. <https://doi.org/10.25077/teknosi.v7i2.2021.108-118>
- Kementerian, P., Ri, K., & Nur, M. (2020). *Indonesian treasury review*. 5, 217–234.
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text Mining in Organizational Research. In *Organizational Research Methods* (Vol. 21, Issue 3). <https://doi.org/10.1177/1094428117722619>
- Komentar, K., Er, T. W. I. T. T., Pengesahan, T. E., Menggunakan, U., & Ma, A. (2012). *Supervised Untuk Klasifikasi*. 16(ISSN 1858). https://www.academia.edu/download/30083202/5-link_vol_16_february_2012.pdf
- Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32–37. <http://ejournal.uin-suska.ac.id/index.php/coreit/article/view/12945>
- Kumar Abhishek, Anshul Khairwa, Tej Pratap, S. P. (2012). A Stock Market Prediction Model using Artificial Neural Network. *IEEE-20180, July*.
- Kurniawan, F., & Wibawa, A. P. (2021). *Text Mining Techniques for Identify Islamophobic Conversation Language by Selecting Preprocessing Feature*. 1–9. <https://www.researchsquare.com/article/rs-1105114/latest.pdf>
- Maitri, A. L., & Sutopo, J. (2019). Rancangan Bangun Chatbot Sebagai Pusat Informasi Lembaga Kursus Dan Pelatihan Menggunakan Pendekatan Natural Language Processing. *Eprints.Uty.Ac.Id*, 1–9. <http://eprints.uty.ac.id/>
- Munasatya, N., & Novianto, S. (2020). Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron. *Techno.Com*, 19(3), 237–244. <https://doi.org/10.33633/tc.v19i3.3630>
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik*, 2(2), 43. <https://doi.org/10.36595/jire.v2i2.117>
- Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Niave Bayes Classifier. *Prosiding*

- Rahman, A., Rapsanjani, W., Junianto, E., Sanjaya, A. R., Sanjaya, A. R., Negasi, C., Network, P. N., & Analysis, S. (2021). *IMPLEMENTASI PROBABILISTIC NEURAL NETWORK DAN WORD EMBEDDING UNTUK ANALISIS*. 3(2), 233–242.
- Rasywir, E., Sinaga, R., & Pratama, Y. (2020). *Evaluasi Pembangunan Sistem Pakar Penyakit Tanaman Sawit dengan Metode Deep Neural Network (DNN)*. 4, 1206–1215. <https://doi.org/10.30865/mib.v4i4.2518>
- Ridok, A., & Latifah, R. (2015). Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN. *Konferensi Nasional Sistem Dan Informatika 2015, Oktober*, 222–227.
- Sartini. (2020). *Analisis Sentimen Twitter Bahasa Indonesia Menggunakan Algoritma Convolutional Neural Network*.
- Septiani, R., Citra, I. P. A., & Nugraha, A. S. A. (2019). Perbandingan Metode Supervised Classification dan Unsupervised Classification terhadap Penutup Lahan di Kabupaten Buleleng. *Jurnal Geografi : Media Informasi Pengembangan Dan Profesi Kegeografian*, 16(2), 90–96. <https://doi.org/10.15294/jg.v16i2.19777>
- Siregar, A. M. (2018). Klasifikasi Algoritma Tf Dan Neural Network Dalam Sentimen Analisis. *AIMS: Jurnal Accounting Information System*, 1(2), 1–8. <https://doi.org/10.32627/aims.v1i2.17>
- Sopian, A., Wiyatno Agus, & Riyandi, A. (2019). Komparasi Algoritma Support Vector Machines Dengan Algoritma Artificial Neural Network Untuk Memprediksi Nilai Persetujuan Kredit Modal Kerja Yang Diberikan. *Jurnal Teknologi Informatika & Komputer*, 5(1), 88–95.
- World Health Organization (WHO). (2020). Novel Coronavirus - Situation Report 22. *World Health Organization*, February, 2019. <https://www.who.int/indonesia/news/novel-coronavirus/qa-for-public>

LAMPIRAN

Source code Cleaning dan Case Folding

```
import re
from bs4 import BeautifulSoup # library untuk menghapus adanya
html dan tanda baca

def case_folding(sosmed):
    replacements = [(r"(http://\www\.|https://\www\.|http://\
|https://\|/)?[a-z0-9]+([\-\.\_]{1}[a-z0-9]+)*\.[a-z]{2,5}(:[0-
9]{1,5})?(\/.*)?$", ""),
                    (r'#[\w]*', ""),
                    (r'@[\w]*', ""),
                    ]
    for old, new in replacements:
        sosmed = re.sub(old, new, sosmed)

    sosmed = BeautifulSoup(sosmed, "lxml").text #hapus html fo
rmatting (&, &lt;, &gt;)
    sosmed = re.sub('[^a-zA-
Z]', ' ', sosmed) #lowercase dan hapus karakter lain
    sosmed = sosmed.lower()
    return sosmed

case_folded_komen = [case_folding(sosmed) for sosmed in sosmed
['Komentar']]
```

Source Code Tokenizing

```
tokenized_komen = [sosmed.split() for sosmed in case_folded_ko
men] # tokenizing dengan split untuk memecah kalimat menjadi k
ata
```

Source Code Stemming

```
pip install PySastrawi #library sastrawi untuk proses stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

join_sosmed = ["".join(sosmed) for sosmed in case_folded_komen
]
stemmed_komen = [stemmer.stem(word) for word in join_sosmed]
```

Source Code TF-IDF

```
import pandas as pd
X = preprocessed_dataset['Komentar']
y = preprocessed_dataset['Kategori']
from sklearn.feature_extraction.text import TfidfVectorizer #library tfidf untuk pembobotan

vectorizer = TfidfVectorizer()
tfidf = vectorizer.fit_transform(X).toarray()
import pickle #library untuk menyimpan model

pickle.dump(vectorizer.vocabulary_, open("feature.pkl", "wb")) #
    menyimpan model tf idf
tfidfDataFrame = pd.DataFrame(tfidf, columns=vectorizer.get_feature_names())
tfidfDataFrame.to_csv('tfidf.csv') #menyimpan hasil tfidf
```

Source Code Pembagian Data

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(tfidf, y,
    test_size=0.1, random_state=0)
```

Source Code Klasifikasi Metode Neural Network

```
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import classification_report, confusion_matrix, recall_score, precision_score, f1_score

score_01= []
f1_score_01 = []
precision_score_01 = []
recall_score_01 = []
cf = []
for i in range(1):
    clf = MLPClassifier(activation='logistic',
        learning_rate_init=0.01,
        tol= -1,
        max_iter=500,
        solver='adam',
        momentum=0,
        learning_rate='constant',
        hidden_layer_sizes=(19))
```

```

clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
print('Confusion matrix:')
confusion_matrix(y_test, y_pred)
import seaborn as sns
import matplotlib.pyplot as plt
f, ax = plt.subplots(figsize=(8,5))
sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, f
mt=".0f", ax=ax)
plt.xlabel("y_head")
plt.ylabel("y_true")
plt.show()

```

Source Code Penyimpanan Model Neural Network

```

import pickle

filename = 'finalized_classifier.pkl'
pickle.dump(clf, open(filename, 'wb'))
loaded_model = pickle.load(open('finalized_classifier.pkl', 'r
b'))
result = loaded_model.score(X_test, y_test)
print(result)

```

Source Code Eksperimen Klasifikasi

```

load_tfidf = pickle.load(open("feature.pkl", 'rb'))

# Create new tfidfVectorizer with old vocabulary
tfidf_new = TfidfVectorizer(vocabulary = load_tfidf)
def predict_new_text(text):
    case_folded_new = case_folding(text)
    tokenized_new = case_folded_new.split()
    stemmed_new = [stemmer.stem(word) for word in tokenized_ne
w]

    cleaned_new_data =[" ".join(stemmed_new)]
    X_tf1 = tfidf_new.fit_transform(cleaned_new_data)
    predict = loaded_model.predict_proba(X_tf1)
    if predict[0][0] > predict[0][1]:

```

```
        return "Hasil Klasifikasi Negatif %.2f%%" % (predict[0
][0]*100)
    else:
        return "Hasil Klasifikasi Positif %.2f%%" % (predict[0
][1]*100)
```