

**SKRIPSI**

**IMPLEMENTASI *LEXICON BASED* DAN ALGORITMA *LONG SHORT-TERM MEMORY* (LSTM) PADA ANALISIS SENTIMEN TWITTER TERHADAP ISU KESEHATAN MENTAL**

**Disusun dan diajukan oleh:**

**YOHANES KASI  
D421 16 019**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2023**

**LEMBAR PENGESAHAN SKRIPSI**

**IMPLEMENTASI LEXICON BASED DAN ALGORITMA LONG SHORT-TERM MEMORY (LSTM) PADA ANALISIS SENTIMEN TWITTER TERHADAP ISU KESEHATAN MENTAL**

**Disusun dan diajukan oleh**

**YOHANES KASI**

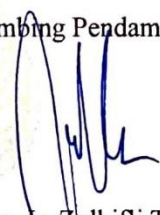
**D42116019**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 11 Agustus 2023 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Elly Warni, ST., MT.  
Nip. 198202162008122001

Dr. Eng. Ir. Zulkifli Tahir, ST., M.Sc.  
Nip. 198404032010121004

Ketua Program Studi,



Prof. Dr. Ir. Indrabayus S.T., M.T., M.Bus.Sys., IPM., ASEAN.Eng  
Nip. 19750716 200212 1 004

## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;  
Nama : Yohanes Kasi  
NIM : D42116019  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Implementasi *Lexicon Based* Dan Algoritma *Long Short-Term Memory (LSTM)*  
Pada Analisis Sentimen Twitter Terhadap Isu Kesehatan Mental

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 04 Agustus 2023

Yang Menyatakan



Yohanes Kasi

## ABSTRAK

**YOHANES KASI.** *Implementasi Lexicon Based dan Algoritma Long Short-Term Memory (LSTM) pada Analisis Sentimen Twitter Terhadap Isu Kesehatan Mental* (dibimbing oleh Elly Warni, S.T., M.T. dan Dr. Eng. Zulkifli Tahir, S.T., M.Sc.).

Pada Oktober 2021 lalu, Direktorat Jenderal Pencegahan dan Pengendalian Penyakit Kementerian Kesehatan menyatakan bahwa pandemi COVID-19 menyebabkan peningkatan pada kasus gangguan jiwa dan depresi hingga 6,5% di Indonesia. Survei yang dilakukan Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI) pada tahun 2020 menemukan, sebanyak 63 persen responden mengalami cemas dan 66 persen responden mengalami depresi akibat pandemi COVID-19. Pengetahuan dan sikap dalam penanganan yang tepat merupakan peran penting dalam menangani isu mental *health*. Berbagai opini masyarakat tentang isu mental *health* seringkali mereka tuangkan lewat media sosial, salah satunya media sosial twitter. Tidak bisa dipungkiri penyampaian pesan dalam media sosial terutama di twitter akan sangat cepat mendapatkan feedback dari khalayak dan tentu saja menimbulkan opini publik, sehingga dapat dibuat analisis sentimen tentang opini masyarakat terhadap isu mental health di Indonesia. Penelitian ini dilakukan untuk melihat kecenderungan sentimen, apakah sentiment tersebut cenderung positif atau negatif dengan menganalisis *tweets* dari pengguna Twitter mengenai mental *health*. Metode yang digunakan pada penelitian ini adalah *Lexicon Based* yang bertujuan untuk mengambil kata-kata yang mengandung sentiment dan *Long Short-Term Memory* (LSTM) sebagai algoritma untuk melakukan klasifikasi sentiment. Hasil penelitian menunjukkan bahwa akurasi yang dihasilkan oleh algoritma LSTM adalah 0.95 atau 95% dengan nilai presisi yang baik pada semua kelas yaitu negatif, netral dan positif secara berurutan sebesar 1.00, 0.92 dan 0.97. Nilai recall juga menunjukkan hasil yang baik untuk kelas netral dan positif masing-masing yaitu 1.00 dan 0.97 sementara pada kelas negatif kurang baik yaitu 0.22. Pada nilai F1-Score untuk kelas negatif, netral dan positif masing-masing yaitu 0.36, 0.96 dan 0.97. Hal ini menunjukkan bahwa algoritma LSTM (*Long Short-Term Memory*) sangat baik untuk mengklasifikasikan sentimen.

Kata Kunci: *Long Short-Term Memory*, LSTM, *Mental Health*, *Deep Learning*

## ABSTRACT

**YOHANES KASI.** *Implementation of Lexicon Based and Long Short-Term Memory (LSTM) Algorithms in Twitter Sentiment Analysis of Mental Health Issues* (supervised by Elly Warni, S.T., M.T. dan Dr. Eng. Zulkifli Tahir, S.T., M.Sc.).

In October 2021, the Directorate General of Disease Prevention and Control of the Ministry of Health stated that the COVID-19 pandemic had caused an increase in cases of mental disorders and depression by up to 6.5% in Indonesia. A survey conducted by the Indonesian Psychiatric Association (PDSKJI) in 2020 found that 63 percent of respondents experienced anxiety and 66 percent of respondents experienced depression due to the COVID-19 pandemic. Knowledge and attitude in proper handling is an important role in dealing with mental health issues. They often express various public opinions about mental health issues through social media, one of which is Twitter. It is undeniable that the delivery of messages on social media, especially on Twitter, will very quickly get feedback from the audience and of course generate public opinion, so that sentiment analysis can be made about public opinion on mental health issues in Indonesia. This research was conducted to see the tendency of sentiment, whether the sentiment tends to be positive or negative by analyzing tweets from Twitter users regarding mental health. The method used in this study is Lexicon Based which aims to retrieve words that contain sentiments and Long Short-Term Memory (LSTM) as an algorithm for classifying sentiments. The results showed that the accuracy produced by the LSTM algorithm was 0.95 or 95% with good precision values for all classes namely negative, neutral and positive respectively 1.00, 0.92 and 0.97. The recall value also showed good results for the neutral and positive class, respectively, namely 1.00 and 0.97, while the negative class was not good, namely 0.22. The F1-Score values for the negative, neutral and positive classes are 0.36, 0.96 and 0.97 respectively. This shows that the LSTM (Long Short-Term Memory) algorithm is very good for classifying sentiments.

Keywords: Long Short-Term Memory, LSTM, Mental Health, Deep Learning

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR .....	vi
DAFTAR TABEL.....	vii
DAFTAR SINGKATAN DAN ARTI SIMBOL .....	viii
DAFTAR LAMPIRAN.....	ix
KATA PENGANTAR .....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian/Perancangan.....	3
1.4 Manfaat Penelitian/Perancangan.....	3
1.5 Ruang Lingkup/Asumsi perancangan .....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Analisis Sentimen .....	4
2.2 Media Sosial.....	4
2.3 Twitter.....	6
2.4 Kesehatan Mental.....	6
2.5 Lexicon Based.....	7
2.6 Long Short-Term Memory Network (LSTM) .....	7
2.7 Confusion Matrix .....	7
BAB 3 METODE PENELITIAN/PERANCANGAN .....	10
3.1 Lokasi Penelitian.....	10
3.2 Tahapan Penelitian.....	11
3.3 Instrumen Penelitian .....	12
3.4 Rancangan Penelitian.....	13
3.5 Teknik Pengumpulan Data.....	13
3.6 <i>Preprocessing Data</i> .....	14
3.7 <i>Lexicon Based</i> .....	17
3.8 Long Short-Term Memory Network (LSTM) .....	19
BAB 4. HASIL DAN PEMBAHASAN .....	21
4.1 Pengumpulan Data .....	21
4.2 Preprocessing Data.....	21
4.3 Pelabelan Data Menggunakan <i>Lexicon Based</i> .....	27
4.4 Analisis Sentiment dengan LSTM ( <i>Long Short-Term Memory</i> ).....	28
4.5 Hasil Analisis .....	32
BAB 5. KESIMPULAN DAN SARAN .....	34
5.1 Kesimpulan .....	34
5.2 Saran.....	34
DAFTAR PUSTAKA .....	35
LAMPIRAN.....	37

## DAFTAR GAMBAR

Gambar 1 Perkembangan Pengguna Media Sosial di Indonesia.....	5
Gambar 2 Tren Pengguna Internet dan Media Sosial di Indonesia.....	5
Gambar 3 Confusion Matrix .....	8
Gambar 4 Lokasi penelitian pada Universitas Hasanuddin Kampus Fakultas Teknik Gowa .....	10
Gambar 5 Alur Tahapan Penelitian.....	11
Gambar 6 Rancangan Sistem .....	13
Gambar 7 Pelabelan menggunakan <i>Lexicon-Based</i> .....	17
Gambar 8 Proses <i>Long Short-Term Memory</i> (LSTM) .....	19
Gambar 9 Contoh data hasil Scrape .....	21
Gambar 10 Grafik Pengujian Model LSTM .....	30
Gambar 11 <i>Confusion Matrix</i> LSTM ( <i>Long Short-Term Memory</i> ) .....	31
Gambar 12 Jumlah Tweet pada tiap Kelas.....	33
Gambar 13 Grafik Jumlah Sentimen Tweet Berdasarkan Bulan .....	33

## DAFTAR TABEL

Tabel 1 Contoh Daftar Kata dan Nilai Polaritas pada <i>InSet Lexicon</i> .....	17
Tabel 2 <i>Tweet</i> Sebelum proses <i>cleaning</i> .....	21
Tabel 3 <i>Tweet</i> Setelah proses <i>cleaning</i> .....	22
Tabel 4 <i>Tweet</i> Sebelum proses <i>case folding</i> .....	22
Tabel 5 <i>Tweet</i> Setelah proses <i>case folding</i> .....	23
Tabel 6 <i>Tweet</i> Sebelum proses <i>tokenization</i> .....	23
Tabel 7 <i>Tweet</i> Setelah proses <i>tokenization</i> .....	24
Tabel 8 <i>Tweet</i> Sebelum <i>Stopwords Removal</i> .....	24
Tabel 9 <i>Tweet</i> Setelah proses <i>Stopwords Removal</i> .....	25
Tabel 10 <i>Tweet</i> Sebelum Normalisasi Teks .....	25
Tabel 11 <i>Tweet</i> Setelah proses Normalisasi Teks .....	26
Tabel 12 <i>Tweet</i> Sebelum <i>Stemming</i> .....	26
Tabel 13 <i>Tweet</i> Setelah proses <i>Stemming</i> .....	26
Tabel 14 Sampel Data dengan Sentimen .....	27
Tabel 15 Sampel Data Hasil <i>Tokenizing</i> .....	29
Tabel 16 Evaluasi Model LSTM.....	31



**DAFTAR SINGKATAN DAN ARTI SIMBOL**

---

Lambang/Singkatan	Arti dan Keterangan
LSTM	<i>Long Short-Term Memory</i>
CSV	<i>Comma Separated Value</i>
NLTK	<i>Natural Language Tool Kit</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>

---

## DAFTAR LAMPIRAN

Lampiran 1 Source Code Preprocessing .....	37
Lampiran 2 Source Code <i>Lexicon Based</i> .....	40
Lampiran 3 Source Code LSTM.....	44
Lampiran 4 Lembar Perbaikan Skripsi .....	48

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya sehingga skripsi yang berjudul “**Implementasi Lexicon Based Dan Algoritma Long Short-Term Memory (LSTM) Pada Analisis Sentimen Twitter Terhadap Isu Kesehatan Mental**” ini dapat terselesaikan dengan baik sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari bahwa banyak kendala yang dihadapi dalam menyelesaikan skripsi ini, namun berkat dorongan, dukungan, bimbingan serta motivasi yang diberikan sehingga penyusunan skripsi ini dapat terselesaikan dengan baik. Ucapan terima kasih serta penghargaan yang setinggi-tingginya penulis sampaikan kepada:

1. Tuhan Yang Maha Esa yang melalui berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Kedua Orang tua penulis, Bapak Fransiskus Ragung dan Ibu Nolana Nona Ros, yang selalu memberikan dukungan, doa, semangat dan kasih sayang serta selalu sabar dalam mendidik penulis sejak kecil.
3. Ibu Elly Warni, S.T., M.T. selaku pembimbing I dan Bapak Dr. Eng. Zulkifli Tahir, S.T., M.Sc. selaku pembimbing II yang selalu menyediakan waktu, tenaga, pikiran dan memberikan bimbingan dalam penyusunan skripsi ini.
4. Bapak Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng., selaku ketua Prodi Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas ilmu, bimbingan dan arahnya selama masa perkuliahan penulis.
5. Segenap Dosen dan Staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu penulis selama masa perkuliahan.
6. Teman-teman igniter16 atas dukungan dan semangat yang telah diberikan.
7. Serta berbagai pihak atas segala dukungan dan bantuannya yang tidak dapat penulis tuliskan satu persatu.

Akhir kata, penulis berharap semoga Tuhan Yang Maha Esa membalas segala kebaikan dari semua pihak yang telah membantu penulis dalam penyusunan skripsi ini, dan semoga skripsi ini dapat memberikan hal yang bermanfaat serta menambah wawasan ilmu untuk pembaca dan juga bagi penulis sendiri.

Makassar, Juli 2023

Penulis

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Pada Oktober 2021 lalu, Direktorat Jenderal Pencegahan dan Pengendalian Penyakit Kementerian Kesehatan menyatakan bahwa pandemi COVID-19 menyebabkan peningkatan pada kasus gangguan jiwa dan depresi hingga 6,5% di Indonesia. Survei yang dilakukan Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI) pada tahun 2020 menemukan, sebanyak 63 persen responden mengalami cemas dan 66 persen responden mengalami depresi akibat pandemi COVID-19.

Dari data tersebut perlu adanya penanganan yang tepat dalam masyarakat, apalagi pandemi membuat perekonomian masyarakat memburuk seperti banyak usaha yang tutup dan pengurangan karyawan, yang secara langsung berakibat pada perekonomian namun juga mental dalam menghadapi segala situasi dimasa pandemi.

Pengetahuan dan sikap dalam penanganan yang tepat merupakan peran penting dalam menangani isu mental *health*. Berbagai opini masyarakat tentang isu mental *health* seringkali mereka tuangkan lewat media sosial, salah satunya media sosial twitter. Tidak bisa dipungkiri penyampaian pesan dalam media sosial terutama di twitter akan sangat cepat mendapatkan *feedback* dari khalayak dan tentu saja menimbulkan opini *public*, sehingga dapat dibuat analisis sentimen tentang opini masyarakat terhadap isu mental *health* di Indonesia.

Pada penelitian ini akan dilakukan untuk melihat kecenderungan sentimen, apakah sentiment tersebut cenderung positif atau negatif dengan menganalisis *tweets* dari pengguna Twitter mengenai mental *health*. Metode yang akan digunakan dalam melakukan analisis sentiment kali ini, yang pertama *Lexicon Based* yang bertujuan mengambil kata-kata yang mengandung sentiment, dan yang kedua proses klasifikasi dengan Algoritma *Long Short-Term Memory* (LSTM) pada penentuan klasifikasi sentimen positif, negatif maupun netral dalam analisis terhadap isu kesehatan mental.

Penggunaan *Lexicon Based* dilakukan karena pemrosesan data menjadi lebih cepat dalam menyaring kata-kata bersentimen. Hal itu berdasarkan dari hasil penelitian sebelumnya oleh Abdul Azis Adjie Sumanjaya, Indriati, Achmad Ridok (2022) yang mengkombinasikan *lexicon based* dan naïve bayes dengan hasil yang mendapatkan akurasi sebesar 0,75, *precision* 0,76, *recall* 0,76, dan *f-measure* 0,75. Penelitian lain dengan metode yang sama juga dilakukan oleh Liza Wikarsa, Apriandy Angdresey, Juan D. Kapantow (2022) dengan menggunakan pendekatan ini, yaitu *dictionary based* dan *corpus-based* pada *lexicon based*, hasil yang mampu dicapai dalam penelitian ini adalah 79,17% akurasi, 89,5% presisi, dan 87,5% *recall*.

Dalam penelitian Nitish Ranjan Bhowmik, Mohammad Arifuzzaman, M. Rubaiyat Hossain Mondal (2022) dengan judul *Sentiment Analysis On Bangla Text Using Extended Lexicon Dictionary And Deep Learning Algorithms*, menemukan bahwa mekanisme berbasis transformator dengan menambahkan LSTM menunjukkan peningkatan yang signifikan yang efektif dalam klasifikasi sentimen. Hasil yang didapatkan adalah *hierarchical attention based LSTM* (HAN-LSTM), *dynamic routing based capsule neural network with Bi-LSTM* (D-CAPSNET-Bi-LSTM) dan *bidirectional encoder representations from Transformers* (BERT) with LSTM (BERT-LSTM) mencapai nilai akurasi masing-masing sebesar 78,52%, 80,82% dan 84,18%.

Berdasarkan uraian diatas, penulis ingin melakukan penelitian analisis sentiment terhadap isu kesehatan mental di Indonesia menggunakan metode *Lexicon Based* dan Algoritma *Long Short-Term Memory* (LSTM).

## 1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian berikut adalah:

1. Bagaimana mengimplementasikan metode *Lexicon Based* dan Algoritma *Long Short-Term Memory* (LSTM) untuk menganalisa sentiment user *tweet*?
2. Bagaiman tingkat akurasi metode *Lexicon Based* dan Algoritma *Long Short-Term Memory* (LSTM) dalam mengklasifikasikan opini, pendapat, keluhan serta saran pengguna twitter yang dijadikan dalam kelas-kelas positif, netral dan negatif?

### 1.3 Tujuan Penelitian/Perancangan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui cara mengimplementasikan metode *Lexicon Based* dan Algoritma *Long Short-Term Memory* (LSTM) dalam menganalisa sentiment user tweet.
2. Mengetahui tingkat akurasi metode *Lexicon Based* dan Algoritma *Long Short-Term Memory* (LSTM) dalam mengklasifikasikan opini, pendapat, keluhan serta saran pengguna twitter yang dijadikan dalam kelas-kelas positif, netral dan negative.

### 1.4 Manfaat Penelitian/Perancangan

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan pengetahuan tentang implementasi *Lexicon Based* sebagai pelabelan kelas sentimen dan Algoritma *Long Short-Term Memory* (LSTM) sebagai penentuan klasifikasi sentimen pada isu kesehatan mental di Indonesia.
2. Membantu masyarakat dan pemerintah dalam melakukan penelitian dari sudut pandang dan opini masyarakat yang tertuang di dalam Twitter sehingga mengetahui tata cara dalam menangani isu kesehatan mental.

### 1.5 Ruang Lingkup/Asumsi perancangan

Yang menjadi ruang lingkup dari penelitian ini adalah sebagai berikut:

1. Penelitian ini melakukan analisa sentimen dengan data yang diambil dari komentar twitter dengan topik isu kesehatan mental masyarakat Indonesia.
2. Hasil akhir dalam penelitian ini adalah kombinasi dari *Lexicon Based* sebagai klasifikasi sentimen dan Algoritma *Long Short-Term Memory* (LSTM) sebagai klasifikasi data.
3. Data hasil opini, pendapat, keluhan serta saran diolah dan diklasifikasikan kedalam kelas-kelas positif, negatif, dan netral.
4. Data twitter yang digunakan membahas tentang isu kesehatan mental masyarakat Indonesia dengan mengambil 1000 data dan dibatasi pada tahun 2021-2022

## **BAB II TINJAUAN PUSTAKA**

### **2.1 Analisis Sentimen**

Analisis sentimen merupakan proses komputasi untuk melakukan klasifikasi terhadap data tekstual sesuai dengan sentimen yang dimiliki. Selain alasan ketersediaan data teks yang melimpah saat ini, penelitian analisis sentimen semakin populer karena alasan kebutuhan dari banyak pihak mengenai pendapat publik terhadap suatu topik tertentu. Proses analisis sentimen dipengaruhi oleh dataset yang digunakan. Untuk dataset yang terdiri dari kumpulan kalimat yang cukup Panjang dan membutuhkan penanganan yang berbeda (Widayat, 2021).

Analisis sentimen biasanya dilakukan untuk mengumpulkan dan mengetahui opini masyarakat dalam postingan Blog, Twitter, Facebook, dan yang lainnya. Analisis sentimen dibutuhkan dengan tujuan untuk mengetahui opini publik terhadap suatu objek. Opini-opini tersebut bisa berupa opini negatif atau positif tergantung dari pandangan publik terhadap objek tersebut (Mas Pintoko & Muslim, 2018). Oleh karena itu dibutuhkan suatu analisis terhadap opini-opini tersebut agar bisa dijadikan tolak ukur dalam menentukan Tindakan kedepannya.

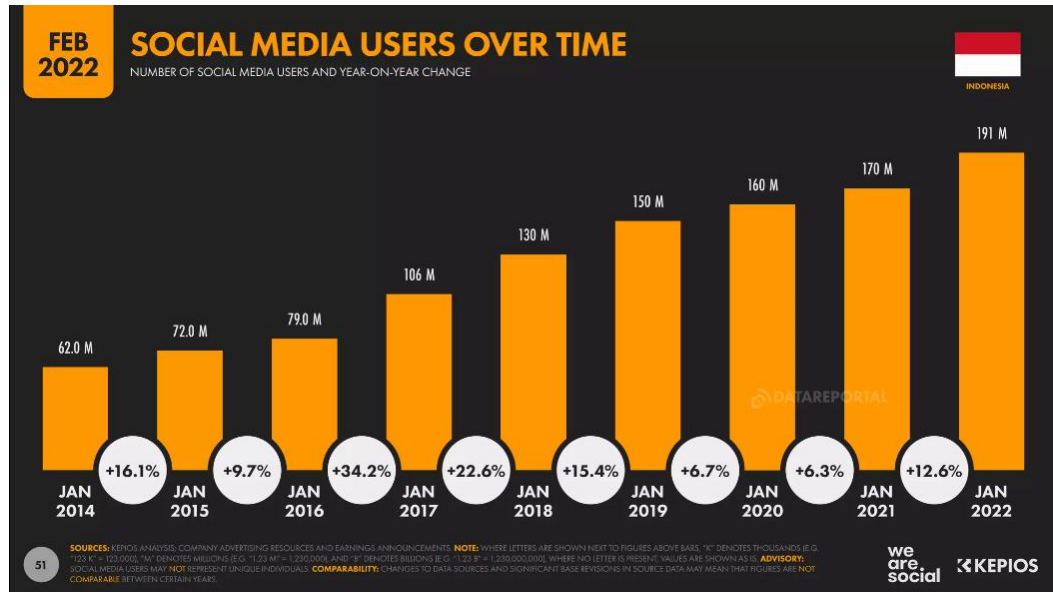
### **2.2 Media Sosial**

Teknologi informasi berkembang pesat dan sangat mempengaruhi hidup manusia. Salah satu teknologi informasi yang paling banyak digunakan masyarakat adalah media sosial. Media sosial merupakan media yang digunakan seseorang atau sekelompok orang dalam melakukan berbagai aktifitas dua arah dalam macam-macam bentuk pertukaran, kolaborasi, dan berinteraksi dalam bentuk tulisan, visual, dan audiovisual (Patricia et al., 2023).

Media sosial sebagai tempat yang dapat mengubah pandangan dan perilaku publik hanya dengan menggunakan tulisan. Setiap tahunnya, pengguna media sosial selalu bertambah dan mudah digunakan terutama di kalangan generasi Y dan Z (Ahmad et al., 2020).

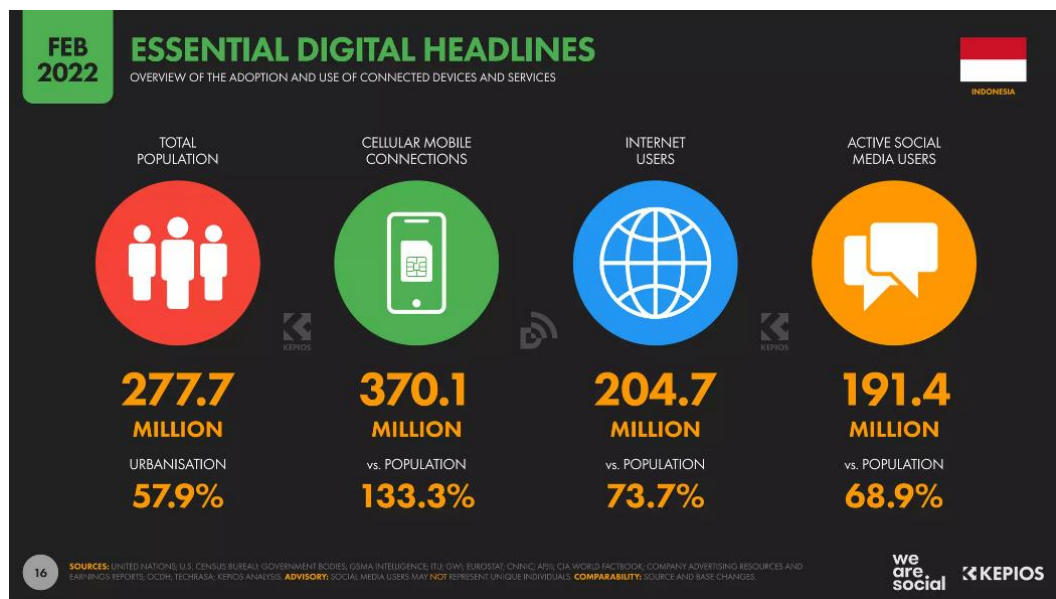
Berikut perkembangan pengguna media sosial di Indonesia (Riyanto, 2022).





Gambar 1 Perkembangan Pengguna Media Sosial di Indonesia

Berikut survei yang dilakukan oleh *We Are School* pada tahun 2022 terkait banyaknya pengguna internet dan media sosial (Riyanto, 2022).



Gambar 2 Tren Pengguna Internet dan Media Sosial di Indonesia

Berdasarkan gambar di atas, berkembang tiap tahun pengguna media sosial begitu pesat dan juga terdapat 204,7 juta pengguna internet, 191,4 juta pengguna media sosial dan 370,1 juta pengguna kartu simcard dari 277,7 juta jumlah populasi penduduk di Indonesia (Riyanto, 2022).

## 2.3 Twitter

Twitter adalah layanan jejaring sosial yang memungkinkan pengguna untuk mengirim, membaca, dan membalas pesan berjenis teks hingga berjumlah 280 karakter. Hal tersebut menyebabkan penggunaannya menjadi mudah guna saling bertukar informasi sehingga setiap individu di planet ini dapat saling terhubung (Madcoms, 2010).

Sejak diluncurkan, twitter telah menjadi salah satu dari situs yang banyak digunakan di kehidupan sehari-hari, dan dijuluki dengan pesan singkat dari Internet. Beberapa istilah dalam Twitter antara lain (Abdul Malik Zuhdi, Ema Utami, Suwanto Raharjo, 2019):

1. Mention, menyebut atau memanggil pengguna Twitter lain dalam sebuah *tweet*, dengan menuliskan '@' diikuti dengan nama pengguna lain.
2. Hashtag, menandai sebuah topik pembicaraan di Twitter, dimulai dengan tanda '#'.
3. *Emoticon*, ekspresi wajah yang direpresentasikan dengan kombinasi antara huruf, tanda baca dan angka.
4. *Trending Topics*, kumpulan dari topik pembicaraan yang sangat populer di Twitter.

## 2.4 Kesehatan Mental

Kesehatan mental yang baik adalah kondisi ketika batin kita berada dalam keadaan tenang dan tentram, sehingga memungkinkan kita untuk menikmati kehidupan sehari-hari dan menghargai orang lain di sekitar. Seseorang yang bermental sehat dapat menggunakan kemampuan atau potensi dirinya secara maksimal dalam menghadapi tantangan hidup, serta menjalin hubungan positif dengan orang lain. Sebaliknya, orang yang kesehatan mentalnya terganggu akan mengalami gangguan suasana hati, kemampuan berpikir, serta kendali emosi yang pada akhirnya bisa mengarah pada perilaku buruk. Penyakit mental dapat menyebabkan masalah dalam kehidupan sehari-hari, tidak hanya dapat merusak interaksi atau hubungan dengan orang lain, namun juga dapat menurunkan prestasi di sekolah dan produktivitas kerja. Oleh sebab itu, sudah saatnya kita menjalankan pola hidup sehat (Promkes Kemkes, 2018).

## 2.5 Lexicon Based

*Lexicon Based* adalah suatu proses pemilihan kata pada dokumen berdasarkan suatu kamus atau *lexicon* yang sudah ada. Metode ini mencocokkan kata-kata yang terdapat dalam kamus *Lexicon* Bahasa Indonesia. Jika *tweets* memiliki kata positif, maka akan digolongkan pada sentimen *tweet* positif. Jika *tweets* memiliki kata negatif, maka akan digolongkan pada sentimen *tweet* negatif. Namun pada kasus lain jika kedua kata ini bernilai sama, maka digolongkan dalam *tweet* netral (Gusti Nur Aulia & Eka Patriya, 2019).

## 2.6 Long Short-Term Memory Network (LSTM)

*Long Term-Short Memory* atau disebut LSTM adalah suatu jenis model turunan dari *Recurrent Neural Network* (RNN) dimana dilakukan suatu modifikasi pada RNN dengan menambahkan sejumlah *memory cell* yang mampu menyimpan suatu informasi untuk jangka yang cukup lama, sekaligus menghapus informasi yang tidak lagi relevan. LSTM juga melengkapi kekurangan RNN yang tidak dapat memprediksi kata berdasarkan informasi lampau yang disimpan. LSTM lebih efisien dalam memproses, memprediksi, sekaligus mengklasifikasikan data berdasarkan urutan waktu tertentu. Serta diharapkan dapat digunakan sebagai solusi untuk mengatasi terjadinya *vanishing gradient* pada RNN saat memproses data *sequential* yang Panjang (Adi Yahyadi & Fitri Latifah, 2022).

## 2.7 Confusion Matrix

*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Pada dasarnya *confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh *system* dengan hasil klasifikasi. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui. *Confusion matrix* memiliki kombinasi nilai prediksi model yang dibagi menjadi 4 bagian seperti pada gambar di bawah ini (Kuncahyo Setyo Nugroho, 2019):

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 3 Confusion Matrix

Tabel di atas merupakan tabel confusion matrix dengan keterangan sebagai berikut:

1. TP (*True Positive*)  
Jumlah nilai aktual kelas positif dan nilai prediksi kelas positif
2. TN (*True Negative*)  
Jumlah data nilai aktual negatif dan nilai prediksi negatif
3. FP (*False Positive*)  
Jumlah data nilai aktual positif dan nilai prediksi negatif
4. FN (*False Negative*)  
Jumlah data nilai aktual negatif dan nilai prediksi positif

*Accuracy* merupakan perhitungan semua data yang diklasifikasikan baik itu positif maupun negatif dalam pengujian. Nilai akurasi dapat menunjukkan kedekatan antara nilai prediksi *system* dan prediksi manusia (Forest & Bayes, 2021). Adapun cara menghitungnya dapat menggunakan rumus dibawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

*Precision* adalah perbandingan data aktual dengan jumlah data yang diprediksi. Nilai *precision* dihasilkan dari nilai prediksi positif berbanding dengan

jumlah nilai yang positif dan di tuliskan pada persamaan berikut (Dewi Sartika & Imelda Saluza, 2022).

$$Precision = \frac{TP}{TP + FP}$$

*Recall* adalah nilai yang menunjukkan tingkat keberhasilan untuk mengetahui kembali informasi mengenai data positif dan negatif yang benar. *Recall* dihasilkan dari jumlah nilai *true* positif dibanding nilai aktual positif, dan di tuliskan pada persamaan berikut (Nur Adinda Salsabila, 2022):

$$Recall = \frac{TP}{TP + FN}$$

F1-Score merupakan *harmonic mean* dari *precision* dan *recall*. Nilai terbaik F1-Score adalah 1.0 sedangkan nilai terburuknya adalah 0, dan di tuliskan pada persamaan berikut (Dewi Sartika & Imelda Saluza, 2022).

$$F1 - score = \frac{2 (Recall \times Precision)}{(Recall + Precision)}$$