# SKRIPSI

# EVALUASI LEXICON-BASED DAN DEEP LEARNING PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERKAIT PERKEMBANGAN METAVERSE DI INDONESIA

**Disusun dan diajukan oleh:**

**RENDY JUNIARTA TODINGBUA**
**D121181336**


**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS HASANUDDIN**

**GOWA**

**2023**

# LEMBAR PENGESAHAN SKRIPSI

## EVALUASI LEXICON-BASED DAN DEEP LEARNING PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERKAIT PERKEMBANGAN METAVERSE DI INDONESIA

Disusun dan diajukan oleh

**RENDY JUNIARTA TODINGBUA**
**D121181336**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 27 September 2023
dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

| Pembimbing Utama, | Pembimbing Pendamping, |
|---|---|
| Dr. Amil Ahmad Ilham ST., M.IT | Elly Warni, ST., M.T |
| NIP 197310101998021001 | NIP 198202162008122001 |

Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 197507162002121004

# PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;
Nama : Rendy Juniarta Todingbua
NIM : D121181336
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Evaluasi Lexicon-Based dan Deep Learning Pada Analisis Sentimen Pengguna Twitter Terkait Perkembangan Metaverse di Indonesia

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, 11 Juni 2023

Yang Menyatakan


Rendy Juniarta Todingbua

# ABSTRAK

**RENDY JUNIARTA TODINGBUA**. *Evaluasi Lexicon-Based Dan Deep Learning Pada Analisis Sentimen Pengguna Twitter Terkait Perkembangan Metaverse Di Indonesia* (dibimbing oleh Dr. Amil Ahmad Ilham, S.T., M.IT. dan Elly Warni ST, MT)

Konsep baru yang menjadi perhatian publik yaitu *'metaverse'* yang merupakan bagian dari pesatnya perkembangan teknologi di dunia. Membawa dunia virtual ke dunia nyata menjadi hal yang sudah sangat mungkin terjadi. Terlebih dengan munculnya pandemi COVID-19 seakan memacu perkembangan teknologi untuk terus maju. Sejak CEO facebook Mark Zuckerberg mengumumkan bahwa facebook akan mengganti nama menjadi meta membuat publik dunia semakin menyoroti teknologi ini. Mengenai konsep metaverse yang ramai diperbincangkan, tidak menutup kemungkinan akan menimbulkan banyak pro dan kontra di antara masyarakat, ada yang menyambut dengan baik dan ada pula yang khawatir mengenai perkembangan teknologi ini.

Penelitian ini bertujuan untuk melihat bagaimana tanggapan masyarakat Indonesia terhadap *metaverse* serta melakukan evaluasi metode Lexicon-Based dan LSTM. Data yang digunakan diperoleh dari *tweet* pengguna twitter dengan kata kunci *metaverse*.

Penelitian ini mengimplementasikan dua algoritma yaitu Lexicon-Based dan LSTM dengan dua model klasifikasi yaitu Long Short-Term Memory (LSTM) dan pendekatan gabungan antaran Lexicon-Based dan LSTM yang kemudian dibandingkan untuk mendapatkan model dengan hasil performa terbaik yang nantinya digunakan untuk melakukan prediksi pada sentimen *tweet* pengguna twitter.

Pada hasil pengujian yang dilakukan, LSTM merupakan model dengan performa terbaik dibandingkan dengan model lainnya dengan nilai akurasi mencapai 93%. Sedangkan untuk model pendekatan gabungan LSTM dan Lexicon-Based sebesar 91%..

Kata Kunci: Analisis Sentimen, Metaverse, Twitter, Lexicon-Based, Deep Learning

# ABSTRACT

**RENDY JUNIARTA TODINGBUA**. *Lexicon-Based and Deep Learning Evaluation Based on Sentiment Analysis of Twitter User Regarding to Metaverse Development in Indonesia* (supervised by Dr. Amil Ahmad Ilham, S.T., M.IT. and Elly Warni ST, MT)

*A new concept that has caught the public's attention is the 'metaverse', which is part of the rapid development of technology in the world. Bringing the virtual world into the real world has become very possible. Moreover, the emergence of the COVID-19 pandemic seems to spur technological developments to continue to advance. Since facebook CEO Mark Zuckerberg announced that facebook would change its name to meta, the world public has increasingly highlighted this technology. Regarding the concept of the metaverse that is being discussed, it is possible that it will cause a lot of pros and cons among the public, some are welcoming and some are worried about the development of this technology.*

*This research aims to see how Indonesian people respond to the metaverse and evaluate Lexicon-Based and LSTM methods. The data used is obtained from twitter user tweets with the keyword metaverse.*

*This research implements two classification models namely Lexicon-Based, Long Short-Term Memory (LSTM) and a combined approach between Lexicon-Based and LSTM which are then compared to get the model with the best performance results which will be used to predict the sentiment of twitter user tweets.*

*In the test results conducted, LSTM is the best performing model compared to other models with an accuracy value of 93%. As for the combined approach model of LSTM and Lexicon-Based at 91%.*

*Keywords*: *Sentiment Analysis, Metaverse, Twitter, Lexicon-Based, Deep Learning*

# DAFTAR ISI

# DAFTAR GAMBAR

# DAFTAR TABEL

# DAFTAR SINGKATAN DAN ARTI SIMBOL

| Lambang/Singkatan | Arti dan Keterangan |
|---|---|
| LSTM | *Long Short-Term Memory* |
| KDD | *Knowledge Discovery in Database* |
| NLP | *Natural Language Processing* |
| API | *Application Programming Interface* |
| CSV | *Comma Separated Value* |
| NLTK | *Natural Language Tool Kit* |
| KBBI | Kamus Besar Bahasa Indonesia |
| TP | *True Positive* |
| TN | *True Negative* |
| FP | *False Positive* |
| FN | *False Negative* |

# DAFTAR LAMPIRAN

# KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala berkat dan karunia yang diberikan kepada penulis sehingga dapat menyelesaikan tugas akhir dengan judul "Evaluasi Lexicon-Based dan Deep Learning pada Analisis Sentimen Opini Pengguna Twitter Terkait Perkembangan Metaverse di Indonesia" sebagai salah satu syarat dalam menyelesaikan pendidikan jenjang Strata-1 di Departemen Teknik Informatika, Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari bahwa pada proses penyusunan dan penulisan tugas akhir ini, penulis banyak mengalami kendala dan kesulitan. Namun berkat dukungan, bantuan dan bimbingan dari berbagai pihak sehingga kendala-kendala tersebut dapat diatasi. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih banyak kepada:

1. Keluarga penulis, Bapak Rudi Sarwanto dan Adriani Rachmat Todingbua selaku kedua orang tua penulis, Ibu Abigael Todingbua, serta Ardi Novrianugrah Todingbua dan Resky Dwi Rara Todingbua selaku kakak kandung penulis yang senantiasa mendoakan, memfasilitasi, serta memberikan dukungan dan semangat yang tiada hentinya dalam menyelesaikan perkuliahan.
2. Bapak Dr. Amil Ahmad Ilham ST., M.IT. selaku pembimbing I dan Ibu Elly Warni, ST., M.T. selaku pembimbing II yang telah menyediakan waktu, tenaga dan pikiran dalam memberikan arahan dan masukan selama pengerjaan dan penulisan tugas akhir.
3. Segenap dosen dan staf Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu penulis selama masa perkuliahan.
4. Teman terbaik, Mage yang senantiasa membantu dan memberikan semangat kepada penulis selama masa perkuliahan.
5. Teman-teman Prodi Teknik Informatika angkatan 2018 atas dukungan, bantuan serta semangat yang telah diberikan.

6. Rekan-rekan dan semua pihak yang namanya tidak dapat disebutkan satu persatu yang telah membantu dalam penulisan tugas akhir ini.
7. Ucapan terimakasih kepada Penulis karena telah mampu melalui semua proses baik dan buruk, suka dan duka serta semua kerja keras selama masa studi.

Penulis berharap semoga Tuhan yang Maha Esa berkenan membalas segala jasa dan kebaikan pada semua pihak yang telah banyak berkontribusi dalam menyelesaikan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna. Oleh karena itu penulis mengharapkan segala bentuk masukan dan saran yang membangun. Semoga tugas akhir ini dapat memberi kontribusi serta menambah wawasan ilmu bagi para pembaca dan semua pihak.

Makassar, 22 Juli 2023

Penulis

# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Perkembangan teknologi yang begitu pesat telah memberi dampak yang signifikan pada peradaban manusia. Terlebih dengan adanya pandemi COVID-19 seakan semakin memacu perkembangan teknologi. Saat ini, konsep baru dari perkembangan teknologi yang menjadi perhatian publik yaitu: '*Metaverse'*, yang merupakan kombinasi dari berbagai aspek teknologi seperti media sosial, *Augmented Reality* (AR), *Virtual Reality* (VR), mata uang digital (*cryptocurrencies*) dan permainan. Konsep *metaverse* pertama kali muncul pada tahun 1992 pada novel fiksi ilmiah *Snow Crash* oleh novelis Amerika Neal Stephenson. Karakter pada *Snow Crash* adalah avatar dan bekerja dalam realitas 3 dimensi (3D), dan dunia *virtual* di mana orang berinteraksi satu sama lain dengan lingkungannya tanpa fisik, keterbatasan dunia nyata yang disebut *metaverse*. Setelah konsep *metaverse* muncul, upaya dan penelitian ekstensif dilakukan untuk membuat *metaverse* menjadi kenyataan. *Metaverse* mengacu pada dunia bersama virtual 3D di mana semua aktivitas dapat dilakukan dengan bantuan layanan augmented dan *virtual reality*.

CEO facebook Mark Zuckerberg menjadi salah satu pionir dalam pengembangan teknologi *Metaverse* dan secara resmi pada 28 Oktober 2021 mengumumkan bahwa Facebook akan mengubah namanya menjadi *Meta* dan akan melakukan investasi yang signifikan pada pengembangan teknologi *Metaverse*. Namun bukan hanya raksasa teknologi *Facebook* yang fokus pada pengembangan *Metaverse*. Beberapa industri telah mengambil peran untuk turut membangun *Metaverse*, sebut saja rumah mode Italia *Gucci, Epic Games, Coca-Cola, Roblox, Minecraft* milik *Microsoft, Nvidia, Apple dan Clinique* yang menjual token digital sebagai batu loncatan menuju *Metaverse* (CNBC Indonesia, 2021). Beberapa negara juga sedang merencanakan untuk mengembangkan kota di *Metaverse* seperti Korea Selatan, Barbados, Arab Saudi dan Indonesia (Maharani, 2021).

Perusahaan teknologi sangat percaya bahwa *Metaverse* akan menjadi masa depan teknologi.

Di tengah perkembangan *Metaverse*, tentu disertai dengan banyaknya opini dari berbagai sisi misalnya dari sudut pandang pemerintah, pengembang teknologi, industri bisnis, pendidikan, agama dan hukum mengenai peluang dan ancaman dari *Metaverse*. Mengadakan *survey* adalah cara tradisional untuk menilai opini dan sikap publik; namun metode ini memiliki batasan seperti jumlah sampel yang tidak terlalu banyak, pertanyaan terbatas, dan data sampel tidak mewakili sampel populasi yang akurat. Untuk mengatasi batasan tersebut data media sosial telah banyak digunakan untuk menganalisis sudut pandang dan opini publik. Twitter merupakan salah satu media sosial yang paling umum digunakan dengan total 1.3 miliar akun serta 368.4 juta pengguna aktif setiap bulan dan 206 juta pengguna aktif setiap hari di seluruh dunia (Ahlgren, 2023); dan menyajikan diskusi publik secara langsung dengan sudut pandang dan opini pengguna yang berbeda dari seluruh dunia. Data tersebut menunjukkan bahwa *Twitter* tidak hanya menjadi media sosial yang digunakan untuk mengemukakan opini dan berdiskusi melainkan juga digunakan oleh para peneliti untuk menganalisis data tersebut.

Pada penelitian yang dilakukan oleh Pane dan Ramdan (2022) menganalisis sentimen masyarakat terhadap kebijakan PPKM menggunakan data Twitter dengan algoritma LSTM untuk mengklasifikasikan data tweet ke dalam kelas sentimen positif dan negatif. Penelitian tersebut menghasilkan akurasi sebesar 94%. Penelitian yang dilakukan oleh Firdaus et al. (2021) menganalisis sentimen evaluasi masukan siswa menggunakan metode *Lexicon-Based* dan kamus *InSet Lexicon* berhasil mendapatkan akurasi sebesar 90.9%

Maka dari itu penelitian ini melakukan evaluasi terhadap metode *Lexicon-Based* dan LSTM serta menganalisa sentimen pengguna *Twitter* untuk melihat kecenderungan persepsi masyarakat terhadap perkembangan *metaverse* di Indonesia pada media sosial *Twitter*.

## 1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian berikut adalah:

a. Bagaimana membangun dan mengevaluasi *Lexicon-Based* dan *Long Short-Term Memory* (LSTM) pada analisis sentimen pengguna *Twitter* terkait perkembangan teknologi *Metaverse* di Indonesia.
b. Bagaimana tanggapan pengguna *Twitter* terkait perkembangan teknologi *Metaverse* di Indonesia*?*

## 1.3 Tujuan Penelitian/Perancangan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

a. Membangun dan mengembangkan sistem analisis sentimen (positif, negatif dan netral) terkait perkembangan *Metaverse* di Indonesia berdasarkan *tweet* pengguna *Twitter* pada bulan Oktober 2021 sampai Desember 2022.
b. Mengevaluasi metode *Lexicon-Based* dan *Deep Learning* dalam melakukan analisis dan klasifikasi sentimen (positif, negatif dan netral) terkait perkembangan *Metaverse* di Indonesia.
c. Mengetahui dan menganalisis tanggapan pengguna *Twitter* terkait perkembangan *Metaverse* di Indonesia berdasarkan hasil klasifikasi sentimen menggunakan metode *Lexicon-Based* dan *Deep Learning*.

## 1.4 Manfaat Penelitian/Perancangan

Adapun manfaat dari penelitian ini adalah sebagai berikut:

a. Memberikan pengetahuan mengenai tanggapan pengguna *Twitter* terkait perkembangan teknologi *Metaverse* di Indonesia menggunakan *Deep Learning*
b. Memberikan pengetahuan tentang evaluasi *Lexicon-Based* dan *Deep Learning* pada analisis sentimen pengguna *Twitter* terhadap perkembangan *Metaverse* di Indonesia
c. Bagi bidang terkait seperti industri, bisnis, pemerintahan dan pendidikan diharapkan hasil penelitian ini dapat menjadi salah satu faktor

pertimbangan bagaimana menyikapi perkembangan *Metaverse* di Indonesia

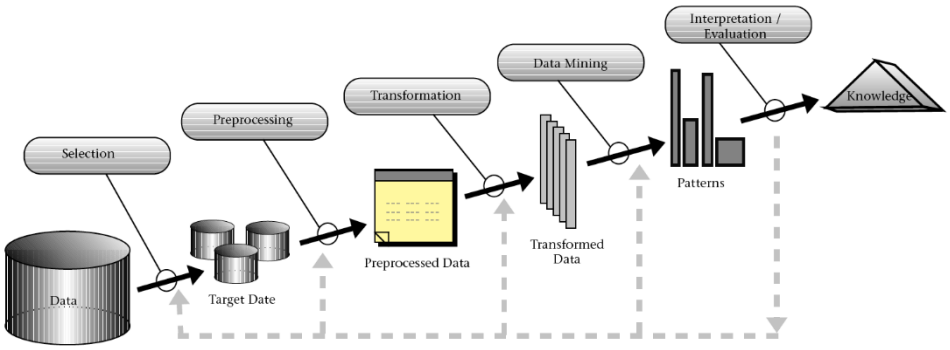
## 1.5 Ruang Lingkup/Asumsi perancangan

a. Data yang digunakan adalah *tweets* yang diperoleh dari *Twitter*
b. Parameter yang digunakan adalah teks *tweet*, waktu *tweet* tersebut dibuat dan lokasi

# BAB II
# TINJAUAN PUSTAKA

## 2.1 *Knowledge Discovery in Database* (KDD)

*Knowledge discovery in database* (KDD) adalah proses untuk mencari dan mengidentifikasi pola atau hubungan yang valid, baru, berpotensi berguna dan pada akhirnya dapat dipahami dalam kumpulan data untuk membuat keputusan penting (Molina-Coronado et al., 2020). *Data science* melibatkan kesimpulan dan iterasi dari banyak hipotesis berbeda. Salah satu aspek kunci dari *data science* adalah proses generalisasi pola dari data. Generalisasi harus berlaku tidak hanya untuk kumpulan data yang digunakan untuk mengamati pola, tapi juga untuk data baru. Proses *knowledge discovery in database* sebagai berikut:




Gambar 2.1.1 Proses KDD

Berdasarkan gambar tersebut, proses *knowledge discovery in database* (KDD) dapat dijabarkan sebagai berikut:

***a. Data Selection***

*Data selection* adalah proses pemilihan data yang dianggap relevan setelah itu data disimpan dalam berkas tersendiri. Data terpilih digunakan dalam proses *data mining*, dimana dipilih data mana yang dibutuhkan untuk diproses lebih lanjut (Li et al., 2018).

**b. *Data Processing***

*Data processing* melibatkan ekstraksi yang valid serta fungsi terstruktur untuk ekstraksi data. Hal ini juga termasuk menerapkan metode pembersihan data untuk menghindari faktor seperti *noise, outlier* atau nilai yang hilang serta penurunan dari fitur baru untuk mengatur data dengan cara yang sesuai untuk penerapan algoritma (Molina-Coronado et al., 2020).

**c. *Data Transformation***

*Data transformation* adalah proses mengubah atau menggabungkan data ke dalam format yang dapat diterima oleh algoritma *data mining* sehingga pola yang ditemukan dapat lebih mudah dipahami (Han et al., 2012). Strategi untuk transformasi data adalah sebagai berikut:

- *Smoothing*, berfungsi untuk menghilangkan *noise* dari data. Teknik operasinya bernama *binning*, regresi dan *clustering*.
- *Attribute Construction* bekerja dengan cara membuat atribut baru yang ditambahkan dari kumpulan atribut yang diberikan untuk membantu proses *data mining*.
- *Aggregation* melakukan operasi agregasi atau ringkasan.
- *Normalization* menskalakan data sehingga berada dalam rentang yang lebih kecil. Ada beberapa metode normalisasi data yaitu *StandardScaler* dan *MinMaxScaler*. *StandardScaler* membuat rata-rata = 0 dan menskalakan data ke varians unit. Sedangkan *MinMaxScaler* menskalakan dalam rentang [-1,1] atau [0,1]. Formula perhitungan normalisasi data menggunakan *MinMaxScaler* ditunjukkan pada persamaan berikut:

  $$x_i = \left(\frac{x_i - x\,min}{x - x\,min}\right)$$

  Keterangan:

  $x_i'$ = hasil normalisasi data ke-i

  $x_i$ = nilai data ke-i

  $x_{min}$ = nilai data minimum dari keseluruhan

  $x_{max}$ = nilai data maksimum dari keseluruhan

  Sedangkan denormalisasi data adalah proses pengembalian data ke awal data sebelum dilakukannya normalisasi data untuk

mendapatkan data asli. Proses denormalisasi dilakukan pada hasil akhir atau output. Perhitungan denormalisasi ditunjukkan pada persamaan berikut:

$$x_i = (y\ (max + min\ )$$

Keterangan:

y = Hasil output

min = Data minimum

max = Data maksimum

**d.*Data Mining***

*Data mining* merupakan inti dari proses KDD, tetapi hanya dapat berhasil jika langkah sebelumnya dilakukan dengan benar. Proses ini terdiri dari pemilihan dan penerapan teknik seperti algoritma *Machine Learning* (ML) untuk mengekstraksi hasil dari data dengan menemukan pola dan hubungan antara fitur. Pemilihan algoritma *data mining* bergantung kepada masalah yang hendak diselesaikan serta sifat dari data yang digunakan (Molina-Coronado et al., 2020).

**e. *Interpretation or Evaluation***

Tahap ini menggunakan sebuah matriks, mengukur kinerja dan efektifitas dari algoritma *data mining* yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

## 2.2 Analisis Sentimen

Analisis sentimen adalah sebuah proses penggalian informasi terhadap suatu entitas yang secara otomatis mengidentifikasi subyektivitas entitas tersebut. Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang menjalankan sistem untuk mengenali, menganalisis dan mengekstraksi opini yang dalam bentuk teks. Tujuannya adalah untuk menentukan apakah teks yang didapatkan menyampaikan pendapat positif, negatif atau netral (Dang et al., 2020). Penerapan analisis sentimen telah menyebar ke hampir setiap bidang seperti layanan kesehatan, keuangan, barang konsumen, fenomena di masyarakat hingga peristiwa sosial dan politik. Dalam berbagai bidang yang berkaitan dengan penggunaan dan dampak analisis sentimen, hal ini menghasilkan banyak sekali

penelitian tentang analisis sentimen saat ini. Menggunakannya menjadi tolak ukur untuk pengembangan produk dan layanan yang baik di dunia bisnis (Undap et al., 2021).

## 2.3 *Lexicon-Based*

*Lexicon based* adalah metode yang menggunakan kata yang telah diberi bobot berdasarkan kamus kata atau *lexicon*. Indikator sentimen yang paling penting adalah kata-kata yang mengandung opini. Misalnya optimal, luar biasa, bagus adalah kata yang bersentimen positif, sedangkan jelek, sampah, mengerikan adalah kata yang bersentimen negatif. Selain kata individu, ada juga ekspresi dan idiom seperti tangan dan kaki. Kata-kata dan ekspresi sentimen sangat penting dalam analisa sentimen. Daftar kata dan frasa seperti itu disebut sentimen *lexicon*. Pemberian nilai terhadap setiap kata dilakukan untuk setiap kata yang mengandung sentimen positif maupun sentimen negatif. Kamus kata pada *lexicon based* dapat dibuat secara manual (Undap et al., 2021). Tujuan penggunaan metode *lexicon based* adalah untuk menentukan orientasi sentimen suatu kalimat.

## 2.4 *Deep Learning*

*Deep learning* merupakan sebuah *artificial intelligence* (AI) yang dapat meniru proses kerja otak manusia. Teknologi ini sangat efektif untuk memproses data mentah dalam jumlah yang sangat besar untuk menciptakan pola yang dapat mengambil keputusan. *Deep learning* merupakan bagian dari *machine learning* namun dirancang dengan kemampuan yang jauh lebih kompleks. *Deep learning* mengadaptasi struktur algoritma berlapis yang disebut jaringan saraf tiruan atau *artificial neural network* (ANN). Pada *machine learning* tradisional, fitur didefinisikan dan diekstraksi baik secara manual atau menggunakan metode pemilihan atribut. Namun pada model *deep learning,* fitur dipelajari dan diekstraksi secara otomatis, meningkatkan akurasi dan kinerja model. Berikut beberapa jenis algoritma pada *deep learning*:

**a. *Convolutional Neural Network* (CNN)**

CNN adalah jenis model *neural network* yang memungkinkan untuk mengekstraksi representasi tingkat yang lebih tinggi untuk pemrosesan gambar. Algoritma ini dirancang khusus untuk memproses data piksel dan citra visual.

**b. *Recurrent Neural Network* (RNN)**

*Recurrent Neural Network* merupakan jenis algoritma *deep learning* yang menerapkan pendekatan yang berurutan atau *sequential*. Nama RNN ini berasal dari fakta bahwa algoritma tersebut bekerja secara berulang. Ini berarti bahwa operasi yang sama dilakukan pada setiap elemen dari suatu urutan, dengan *output* yang tergantung pada *input* saat ini dan operasi sebelumnya (Merinda Lestandy et al., 2021). RNN dirancang untuk mendeteksi pola dalam kumpulan data seperti teks, tulisan tangan, kata-kata yang diucapkan dan data numerik dalam bentuk deret waktu seperti data dari sensor dan informasi tentang pergerakan saham.

**c. *Long Short-Term Memory Network* (LSTM)**

LSTM adalah salah satu modifikasi dari *recurrent neural network*. LSTM hadir untuk melengkapi kekurangan RNN yang tidak dapat memprediksi kata berdasarkan data lampau yang disimpan dalam waktu lama. Dengan demikian, LSTM dapat mengingat kumpulan informasi yang sudah lama disimpan dan juga menghapus informasi yang sudah tidak relevan. LSTM lebih efisien dalam memproses, memprediksi sekaligus mengklasifikasikan data berdasarkan urutan waktu tertentu (Pane & Ramdan, 2022).

**d. *Self Organizing Maps* (SOM)**

*Self organizing maps* (MAPS) adalah teknik pada *neural network* yang bertujuan untuk melakukan visualisasi dengan mengurangi jumlah dimensi pada data menggunakan *self organizing neural networks*. Visualisasi data nantinya berguna untuk menyelesaikan masalah yang sulit dipecahkan oleh manusia. SOM diciptakan untuk membantu pengguna agar mudah memahami informasi berdimensi tinggi dengan mudah. Jaringan ini merupakan jenis yang mudah dilatih menggunakan metode *unsupervised learning* (Khazri, 2019).

***f. Transfer-Learning***

*Transfer-learning* adalah suatu metode pada *machine learning* yang menggunakan pengetahuan dari model yang sebelumnya telah dilatih (*domain sumber*) untuk mempercepat dan meningkatkan kinerja model baru yang akan dilatih (*target domain*). Pada konteks *deep learning*, *transfer learning* terbukti sangat efektif dalam mempercepat pelatihan model, meningkatkan akurasi, dan memperluas penerapan model ke berbagai tugas dan domain (Prottasha et al., 2022).

## 2.5 *Metaverse*

*Metaverse* adalah dunia virtual, lingkungan multi-pengguna yang menyatukan antara realitas fisik dan dunia virtual. Hal ini didasarkan pada konvergensi teknologi yang memungkinkan interaksi multi sensor dengan lingkungan virtual dan objek digital menggunakan konsep *virtual reality* (VR) dan *augmented reality* (AR). Oleh karena itu *metaverse* adalah jejaring sosial yang menghubungkan multi pengguna dalam suatu lingkungan daring secara persisten. Hal ini memungkinkan untuk mengimplementasikan interaksi pengguna secara real-time dan interaksi dinamis dengan objek digital. Iterasi pertama adalah jaringan dunia virtual di mana avatar dapat berinteraksi satu sama lain. Iterasi modern *metaverse* menawarkan platform VR sosial dan teknologi imersif yang kompatibel dengan permainan *online, open world*, dan ruang AR kolaboratif (Mystakidis, 2022).

## 2.6 Twitter

Twitter adalah layanan jejaring sosial atau mikroblog *online* yang memungkinkan pengguna mengirim, membaca, dan membalas pesan teks (juga dikenal sebagai *tweet*) hingga 280 karakter. Awalnya, twitter hanya mengizinkan pengguna untuk mengirim *tweet* maksimal 140 karakter, namun pada 7 november 2017 twitter meningkatkan batas menjadi 280 karakter. Di twitter, pengguna yang tidak terdaftar hanya dapat membaca *tweet* pengguna lain sementara pengguna terdaftar dapat menulis, membagikan dan menyukai *tweet* melalui antarmuka

*website* dan aplikasi *smartphone android* dan IOS (*Iphone*). *Twitter* banyak digunakan sebagai alat kampanye politik, dijadikan sarana protes, pembelajaran hingga media komunikasi darurat (Kwak et al., 2010).

## 2.7 *Performance Metrics*

*Performance metrics* atau matriks performa model adalah seperangkat matriks atau ukuran yang digunakan untuk mengukur kinerja model dalam memprediksi nilai target dari data. *Performance metrics* digunakan untuk mengevaluasi seberapa baik model dapat mempelajari pola dari data yang diberikan dan memprediksi nilai target dengan akurasi yang tinggi. *Performance metrics* dapat berbeda-beda tergantung pada jenis masalah *machine learning* yang sedang dipecahkan. Berikut beberapa jenis *performance metrics*:

***a. Confusion Matrix***

*Confusion matrix* atau *error matrix* memberikan informasi perbandingan antara hasil klasifikasi dari sistem (model) dengan hasil klasifikasi yang sebenarnya. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada sekumpulan data uji yang diketahui nilai sebenarnya. *Confusion matrix* memiliki 4 kombinasi nilai prediksi dan nilai aktual seperti gambar berikut:

| | | Kelas Prediksi | |
|---|---|---|---|
| | | Positive | Negative |
| Kelas Aktual | Positive | TP | FN |
| | Negative | FP | TN |

**(a)**

| | | Kelas Prediksi | | | |
|---|---|---|---|---|---|
| | | $C_1$ | $C_2$ | ... | $C_N$ |
| Kelas Aktual | $C_1$ | $C_{1,1}$ | FP | ... | $C_{1,N}$ |
| | $C_2$ | FN | TP | ... | FN |
| | ... | ... | ... | ... | ... |
| | $C_N$ | $C_{N,1}$ | FP | ... | $C_{N,N}$ |

**(b)**

Gambar 2.7.1 Confusion Matrix: (a) *Binary Classification* (b) *Multiclass Classification* (Markoulidakis et al., 2021)

Terdapat 4 istilah pada *confusion matrix* sebagai representasi hasil proses klasifikasi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) (Nugroho, 2019).

- *True Positive* (TP)

  Merupakan kelas positif yang diprediksi sebagai kelas positif.
- *True Negative* (FN)

  Merupakan kelas negatif yang diprediksi sebagai kelas negatif.
- *False Positive* (FP)

  Merupakan kelas negatif yang diprediksi sebagai kelas positif.
- *False Negative* (FN)

  Merupakan kelas positif yang diprediksi sebagai kelas negatif.

*Confusion Matrix* untuk klasifikasi multi-kelas merupakan sebuah matriks dengan skala N x N dimana N merupakan jumlah kelas. Setiap baris dalam matriks mewakili kelas sebenarnya serta setiap kolom mewakili kelas yang diprediksi.

***b. Accuracy***

*Accuracy* digunakan untuk mengukur seberapa akurat model dalam memprediksi kelas target dari suatu data. Accuracy menghitung rasio prediksi yang benar (positif dan negatif) terhadap total keseluruhan data yang dituliskan seperti persamaan berikut:

$$accuracy = \frac{TP + TN}{Jumlah\ Data}$$

Dalam konteks *machine learning*, akurasi digunakan sebagai matriks performa model untuk mengevaluasi seberapa baik model dapat memprediksi kelas target yang benar. Semakin tinggi nilai akurasi, semakin baik kinerja model dalam memprediksi kelas target. Namun akurasi juga dapat memberikan hasil yang bias pada data yang tidak seimbang yaitu ketika jumlah sampel pada kelas yang satu lebih banyak daripada yang lain. Oleh karena itu, selain akurasi, perlu juga digunakan matriks performa model lainnya seperti *precision, recall, F1-Score* atau AUC untuk mengevaluasi kinerja model secara lebih komprehensif dan mendalam.

***c. Precision and Recall***

*Precision* adalah perbandingan antara data aktual dengan jumlah data yang diprediksi dan dituliskan seperti persamaan berikut:

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

Sementara *Recall*, adalah tingkat keberhasilan model melakukan klasifikasi dan dituliskan seperti persamaan berikut:

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

*Precision* dan *recall* memiliki perbedaan pada variabel yang digunakan, *precision* menggunakan variabel *False Positive* (FP) sedangkan *recall* menggunakan variabel *False Negative* (FN). Semakin kecil *False Positive* (FP) maka *precision* semakin besar dan semakin kecil *False Negative* (FN) maka *recall* semakin besar (Setiawan, 2020).

***d. F1-Score***

F1-Score adalah *harmonic mean* dari presisi dan *recall* yang dituliskan seperti persamaan berikut:

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall}\right)$$

Nilai tertinggi *F1-Score* adalah 1.0 dan nilai terendah adalah 0. Jika *F1-Score* memiliki skor yang baik mengindikasikan bahwa model memiliki presisi dan *recall* yang baik (Kanstrén, 2020).

## 2.8 *InSet Lexicon*

*InSet Lexicon* adalah kamus Lexicon bahasa Indonesia yang dibuat untuk mengidentifikasi opini tertulis dan mengklasifikasikannya menjadi opini positif dan negatif yang dapat digunakan untuk analisis sentimen publik terhadap topik, peristiwa atau produk tertentu (Koto & Rahmaningtyas, 2017). Fajri Koto dan Gemal Y. Rahmaningtyas mengembangkan *InSet Lexicon* pada penelitian sebelumnya menggunakan kata-kata yang dikumpulkan dari *Twitter* sebagai media sosial populer di Indonesia. Hasil pengujian dan evaluasi penelitian menunjukkan *InSet Lexicon* dapat memberikan performa yang memuaskan sebagai kamus sentimen bahasa Indonesia dengan akurasi sebesar 90.9% (Firdaus et al., 2021).