

BAB II

LANDASAN TEORI

2.1. PYTHON

Python adalah bahasa pemrograman tingkat tinggi tujuan umum yang banyak digunakan. Ini dibuat oleh Guido van Rossum pada tahun 1991 dan dikembangkan lebih lanjut oleh *Python Software Foundation*. Itu dirancang dengan penekanan pada keterbacaan kode, dan sintaksnya memungkinkan *programmer* untuk mengekspresikan konsep mereka dalam baris kode yang lebih sedikit [6].

Pemrograman bahasa *python* difokuskan untuk digunakan dalam menganalisa data, visualisasi data, membuat dan mengembangkan *AI*. Pemrograman *Python* adalah pemrograman yang paling mudah dipelajari dengan kode pendek dan tidak susah [7].

Banyak programmer dan peneliti yang beralih ke penggunaan bahasa ini. Python memiliki berbagai kegunaan, seperti pengembangan aplikasi web, aplikasi *desktop*, *IoT*, dan banyak lagi. *Python* juga memiliki integrasi dengan sistem *database* dan kemampuan untuk membaca dan mengubah file, sehingga sering digunakan dalam *prototyping* atau pengembangan perangkat lunak dengan cepat dan andal. Perkembangan dan kegunaan *Python* sangat luar biasa sehingga bahasa ini menjadi bahasa pemrograman nomor satu di dunia. Namun, penggunaan *Python* masih belum begitu populer di Indonesia [8].

2.2. MACHINE LEARNING

Artificial intelligence (AI) dan *machine learning* saat ini kembali memasuki fase *booming* setelah beberapa dekade mengalami pasang surut. Kecerdasan buatan kembali digandrungi, dimana penerapannya dilakukan secara massif pada aplikasi-aplikasi bisnis dan social media jaman *now* seperti *Facebook*, *Twitter*, *Google*, *Amazon*, dan bahkan berbagai aplikasi besar dari Indonesia seperti *Gojek*, *Tokopedia*, dan sebagainya. *Machine learning* merupakan salah satu cabang dari ilmu Kecerdasan Buatan, khususnya yang mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya [9].

Machine learning merupakan bidang lain dari ilmu komputer yang merancang sebuah algoritma agar memungkinkan sebuah komputer untuk belajar melalui data sehingga sering dikatakan sebagai *learn from data*. Jadi *machine learning* adalah pemrograman komputer yang menggunakan data masa lalu yang digunakan untuk pembelajaran model sehingga mendapatkan performa yang optimal dalam menggali informasi dari suatu kumpulan data.

Inti *machine learning* adalah untuk membuat model yang merefleksikan pola-pola data. Sedangkan menurut Tom M. Mitchell definisi *machine learning* adalah sebuah program komputer yang belajar dari pengalaman dari tugas yang dibebankan dengan kinerjanya yang terukur. Secara garis besar *machine learning* dibagi menjadi 3 yaitu *Supervised Learning*, *Unsupervised Learning* dan *Reinforcement Learning*.

a. *Supervised Learning* adalah algoritma machine learning yang proses pembelajarannya dibawah pengawasan. Yang termasuk dalam *supervised learning* diantaranya adalah *classification* dan *regression*.

b. *Unsupervised Learning* adalah algoritma machine learning yang proses pembelajarannya tanpa pengawasan. Yang termasuk dalam *unsupervised learning* diantaranya adalah *clustering* dan *dimensionality reduction*.

c. *Reinforcement Learning* adalah algoritma machine learning yang bisa membuat *agent software* mesin bekerja dan bekerja secara otomatis untuk menentukan perilaku ideal sehingga dapat memaksimalkan kinerja algoritma. Yang termasuk dalam *Reinforcement Learning* diantaranya adalah *real-time decision, robot navigation, learning tasks, skills acquisition dam game AI* [10].

Neural network adalah salah satu pendekatan pembelajaran terawasi yang telah mendapatkan popularitas yang cepat dalam beberapa tahun terakhir. Itu karena kinerjanya. Ini adalah jaringan saraf yang sangat khusus. Salah satu faktor yang mempengaruhi performansi atau nilai presisi yang dihasilkan adalah banyaknya iterasi [11].

2.3. NATURAL LANGUAGE PROCESSING (NLP)

Di era sekarang kecerdasan buatan melalui chatbot menjadi salah satu andalan dalam pemecahan berbagai masalah. Salah satu jenis chatbot yang populer adalah *ChatGPT*, yang menggunakan teknologi *NLP (Natural Language Processing)* untuk memahami dan merespons pertanyaan dan permintaan pengguna [12].

Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang melibatkan desain dan implementasi sistem dan algoritma yang mampu berinteraksi melalui bahasa manusia. Berkat kemajuan pembelajaran mendalam terkini, aplikasi *NLP* telah menerima peningkatan kinerja yang belum pernah terjadi sebelumnya [13].

Pemrosesan bahasa alami (*NLP*) baru-baru ini mendapat banyak perhatian untuk mewakili dan menganalisis bahasa manusia secara komputasi. Ini telah menyebarkan aplikasinya di berbagai bidang seperti terjemahan mesin, deteksi spam email, ekstraksi informasi, ringkasan, medis, dan menjawab pertanyaan, dll [14].

2.4. TENSORFLOW

TensorFlow 2.0, dirilis pada Oktober 2019, mengubah kerangka kerja dengan banyak cara berdasarkan masukan pengguna, untuk membuatnya lebih mudah digunakan (misalnya, dengan menggunakan *API Keras* yang relatif sederhana untuk pelatihan model) dan lebih berkinerja. Pelatihan terdistribusi lebih mudah dijalankan berkat *API* baru, dan dukungan untuk *TensorFlow Lite* memungkinkan penerapan model di berbagai platform yang lebih besar. Namun,

kode yang ditulis untuk versi *TensorFlow* sebelumnya harus ditulis ulang terkadang hanya sedikit, terkadang secara signifikan untuk memanfaatkan fitur *TensorFlow 2.0* baru secara maksimal.

TensorFlow memungkinkan *developer* membuat grafik aliran data struktur yang mendeskripsikan bagaimana data bergerak melalui grafik, atau serangkaian node pemrosesan. Setiap node dalam grafik mewakili operasi matematika, dan setiap koneksi atau tepi antar *node* adalah larik data multidimensi, atau *tensor*. *Tensorflow* menggabungkan banyak model dan algoritma *machine learning* yakni *deep learning (neural network)*. *Framework* ini disusun menggunakan *Python front-end API* untuk membuat suatu aplikasi penggunaannya, dan menggunakan *C++* yang memiliki kinerja terbaik dalam hal eksekusi.

Tensorflow dapat melatih dan menjalankan *neural network* untuk keperluan mengklasifikasikan tulisan tangan, pengenalan gambar/objek, serta menggabungkan suatu kata. Selanjutnya *re-current neural network* yang merupakan model *sequential* dapat digunakan untuk *Natural Language Processing (NLP)*. Selain itu, *tensorflow* digunakan pada skala yang besar untuk produksi dengan menggunakan model yang sama ketika proses *training* data. *Tensorflow* ini merupakan *library* yang sangat populer pada kalangan data *enthusiast* terutama pelaku *machine learning*[15].

Tensorflow adalah sebuah Bahasa yang dikembangkan oleh *Google Brain Teams*, dan dirilis pada tahun 2015. *Tensorflow* pada dasarnya adalah sebuah *framework* yang dapat digunakan untuk proses pengembangan dan pembuatan sebuah *AI*.

Kelebihan Tensorflow sendiri memiliki *API* yang mudah digunakan untuk berbagai task dalam proses pembuatan *AI*. *Tensorflow* juga lebih mudah digunakan untuk pembuatan *AI*, *Tensorflow* juga memiliki komunitas besar sehingga memudahkan developer dalam proses pembuatan dan pengembangan. *Tensorflow* juga banya digunakan di tingkat industri.

Kekurangan dari *Tensorflow* adalah *Tensorflow* akan lebih sulit digunakan dan kompleks digunakan jika digunakan di proses pembuatan *AI* yang lebih kompleks. Dan *Tensorflow* memiliki dokumentasi yang tidak lengkap [16].

Tensorflow adalah Pustaka sumber terbuka yang dibuat oleh *Google*, yang bertujuan untuk membantu pengembang dalam membuat berbagai model *machine learning*. Peningkatan terbaru dalam ruang lingkup *deep learning* menciptakan kebutuhan untuk mempermudah dan mempercepat dalam membangun sebuah *neural network*. *Tensorflow* merupakan salah satu *library* yang mengatasi masalah ini dengan sangat baik, dengan menyediakan luas berbagai *API* dan alat untuk membantu pengembang fokus pada spesifik masalah mereka, daripada berurusan dengan permasalahan matematika dan masalah skalabilitas [17].

2.5. BIDIRECTIONAL LSTM (Bi-LSTM)

Bi-LSTM adalah variasi dari *Recurrent Neural Network (RNN)* yang dikembangkan untuk melatih model *Neural Network* menggunakan sequential data dengan input masa lalu dan masa depan [18] .

Salah satu varian dari *LSTM* adalah *Bidirectional LSTM (BiLSTM)*. *Input* yang dimasukkan ke dalam *BiLSTM* ada 2 jenis yaitu *input forward* dan *input backward*. *Output* dari lapisan ini umumnya digabungkan menjadi satu. Dengan

layer ini, model dapat mempelajari informasi masa lalu (*past*) dan informasi masa mendatang (*future*) untuk tiap sekuen *input* [19].

Jaringan *Bidirectional LSTM (BiLSTM)* dengan kemampuan mengekstraksi informasi kontekstual dari rangkaian fitur dengan menangani ketergantungan maju dan mundur. Selain itu, *BiLSTM* memungkinkan kami melihat ke depan dengan menggunakan *LSTM* ke depan, yang memproses urutan dalam urutan kronologis, dan *LSTM* ke belakang, yang menangani urutan dalam urutan terbalik. *Outputnya* kemudian merupakan gabungan dari status *LSTM* maju dan mundur yang sesuai [20].

2.6. KLASIFIKASI SENTIMEN

Klasifikasi sentimen merupakan salah satu cabang dari *Text mining*. Klasifikasi sentimen dapat menjadi sesuatu yang penting dalam proses evaluasi terhadap sebuah topik permasalahan. Tujuan utama dari klasifikasi sentimen adalah untuk mencari tahu polaritas dari sentimen positif, negatif dan netral. Klasifikasi sentimen merupakan salah satu cara untuk mengetahui pendapat seseorang atau sekelompok orang terhadap isu, produk, layanan atau golongan tertentu. Klasifikasi sentimen dapat dilakukan dengan mengumpulkan data melalui *Twitter* dengan topik tertentu [21].

Klasifikasi sentimen digunakan untuk melakukan penilaian suatu entitas tekstual yang menunjukkan emosi atau sentimen. Tujuan utama dari klasifikasi sentimen teks adalah mengelompokkan suatu data teks agar ditemukan polaritasnya. Jumlah data yang besar berdampak pada semakin sulitnya

melakukan klasifikasi sentimen. Kesulitan melakukan klasifikasi sentimen pada data yang besar ini dapat diatasi menggunakan *deep learning* [22] .

Analisis sentimen adalah sebuah set algoritma dan Teknik yang digunakan untuk mendeteksi sentimen yang terdapat di suatu teks dan mengklasifikasikannya ke dalam tiga jenis klasifikasi, yaitu positif, negatif, dan netral [23].

2.7. DATASET

Dataset adalah kumpulan data yang digunakan untuk tujuan pembelajaran mesin. Data dapat berupa piksel, gambar, seri waktu, atau jenis data lainnya yang merupakan realisasi dari variabel acak yang diminati. *Dataset* biasanya dilabeli dengan kelas, atribut, atau nilai yang ingin diprediksi oleh model *ML*. *Dataset* harus memiliki kualitas yang baik untuk memastikan kebenaran dan kinerja model *ML* [24].

Dataset adalah sekumpulan data atau himpunan data. Dalam data tabular, Kumpulan data sesuai dengan satu/lebih tabel basis data (*database*), pada setiap kolom tabel mewakili variabel tertentu dan setiap baris sesuai dengan catatan tertentu dari kumpulan data yang bersangkutan. Kumpulan data mencantumkan nilai untuk setiap variabel. Kumpulan data terdiri dari kumpulan file atau dokumen. *Dataset* adalah Kumpulan data yang menggambarkan suatu objek yang memiliki atribut [25].

Dataset adalah objek yang mempresntasikan data dan relasinya di memori, dataset berisi koleksi dari data tabel dan data, jenis dataset ada dua macam yaitu :

a. *Private dataset*

Private dataset adalah data yang didapatkan atau diambil dari organisasi yang dijadikan tempat atau objek penelitian

b. *Public dataset*

Public dataset adalah dataset yang dapat diambil dari *repository public* (penyimpanan umum) yang telah disepakati oleh peneliti untuk dipublikasi kepada umum [26].

Pada penelitian ini *dataset* yang digunakan merupakan dataset publik yang telah diberi label, didapatkan dari *Kaggle*, yakni sebuah *platform online* yang menyediakan berbagai kumpulan data, kode, dan kompetisi untuk para penggiat *data science* dan *machine learning*.

Berikut adalah informasi dari dataset yang digunakan :

Tabel 2.1 Informasi *dataset* [27]

Nama	<i>Sentiment Analysis Dataset</i>
Ukuran	57 MB
Update terbaru	2021
Pemilik	Abhishek Shrivastava
Bahasa	Inggris
Tautan	https://www.kaggle.com/datasets/abhi8923shriv/sentiment-analysis-dataset?resource=download&select=test.csv
Keterangan	Data pelatihan dibuat secara otomatis, dibandingkan dengan manusia yang membuat anotasi tweet secara manual. Dalam pendekatan kami, kami berasumsi bahwa setiap tweet dengan emotikon positif, seperti :), adalah positif, dan tweet dengan emotikon negatif, seperti :(, adalah negatif

2.8. PREPROCESS

Preprocessing adalah sebuah langkah penting dalam proses penambangan data. Data yang akan digunakan dalam proses penambangan data tidak selalu dalam kondisi terbaik untuk diproses. Ada kalanya dalam data tersebut terdapat beberapa masalah yang nantinya dapat mempengaruhi hasil yang diberikan dari proses penambangan itu sendiri seperti terdapat nilai yang hilang, data yang berlebihan, *outliner*, atau format data yang tidak sesuai dengan sistem. Oleh karena itu untuk mengatasi masalah tersebut perlu dilakukan tahap *preprocessing*. *Preprocessing* adalah salah satu langkah dalam menghilangkan masalah yang dapat mengganggu hasil dari pada proses klasifikasi data [28].

Preprocessing adalah suatu proses dimana dataset yang di *collect* dari media sosial *twitter* akan dibersihkan dari elemen-elemen yang tidak dibutuhkan sehingga nantinya akan mendapatkan data yang memiliki kualitas dan sesuai apa yang diinginkan [29].

Berikut adalah beberapa langkah preprocess yang digunakan dalam penelitian ini :

a. Ekstrak kalimat dan label dari *dataset*

kalimat dan label dalam dataset berada pada kolom yang berbeda. Kalimat perlu diambil untuk dilakukan preprocessing berikutnya, sementara label akan digunakan dalam proses pelatihan.

b. *Stop Word Removal*

Stop Word Removal merupakan tahap menghilangkan konjungsi yang sering muncul. Kata-kata ini biasanya ada suatu fungsi tetapi tidak mempunyai arti dan tidak memberi bobot pada suatu pendapat atau kalimat

c. *Stemming*

Stemming merupakan tahap pembersihan afiks (imbuhan kata), baik prefiks, sufiks maupun kombinasi keduanya. Dengan *stemming*, kata-kata yang memiliki kata dasar yang sama akan dianggap memiliki token yang sama. Ini membantu dalam meningkatkan kinerja pemrosesan data.

d. *Cleansing Data*

Cleansing Data adalah proses membersihkan atau menghilangkan sesuatu, seperti tanda baca, angka, URL, dan kata-kata yang dianggap tidak penting.

e. *Case Folding*

Dalam proses ini, kata-kata yang berhuruf besar diubah menjadi huruf kecil. Proses ini dilakukan untuk menghindari duplikasi, yang dibedakan dari huruf besar/kecil (peka huruf besar/kecil)

f. *Tokenizing*

Tokenisasi adalah membagi atau memecah suatu kalimat yang sebelumnya dipisahkan oleh spasi menjadi token kata-kata itu penyusunnya. Tokenisasi ini merupakan proses penting untuk klasifikasi [30] .

2.9. FLOWCHART / DIAGRAM ALUR

Flowchart adalah suatu teknik untuk menggambarkan urutan logika dari suatu prosedur penyelesaian masalah. Dengan kata lain, *flowchart* merupakan langkah – langkah penyelesaian masalah yang disajikan dalam bentuk simbol-simbol tertentu[31].

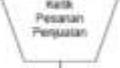
Flowchart adalah penggambaran secara grafik dari tahapan penyelesaian suatu masalah yang terdiri atas sekumpulan simbol, Dimana masing-masing simbol mempresentasikan suatu kegiatan tertentu [32].

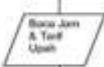
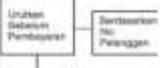
Diagram alur program adalah alat yang sangat berguna dalam pengembangan program. pertama, setiap kesalahan atau kelalaian dapat lebih mudah dideteksi dari diagram alur program daripada bentuk program karena diagram alur program adalah representasi bergambar dari logika suatu program. kedua, *flowchart* program dapat diikuti dengan mudah dan cepat. ketiga, ini berfungsi sebagai jenis dokumentasi, yang mungkin sangat membantu jika diperlukan modifikasi program di masa mendatang.

Lima aturan berikut harus diikuti saat membuat diagram alur program

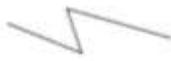
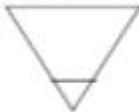
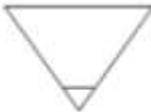
- a. Hanya simbol standar yang harus digunakan dalam diagram alur program.
- b. Logika program harus menggambarkan aliran dari atas ke bawah dan dari kiri ke kanan.
- c. Setiap simbol yang digunakan dalam diagram alur program hanya boleh memuat satu titik masuk dan satu titik keluar, dengan eksepsi simbol keputusan. ini dikenal sebagai aturan tunggal.
- d. Operasi yang ditampilkan dalam simbol [diagram alur program harus dinyatakan secara independen dari bahasa pemrograman tertentu.
- e. Semua cabang pengambilan keputusan harus diberi label yang baik [33].

Salah satu bagan yang digunakan untuk menyatakan algoritma disebut dengan bagan alir (flowchart) [34].

SIMBOL	ARTI	CONTOH	SIMBOL	ARTI	CONTOH
 Dokumen	IO dalam format yang dicetak		 On-line Storage	IO yang menggunakan penyimpanan langsung	
 Magnetic Tape	IO yang menggunakan pita magnetik		 Punched Tape	IO yang menggunakan pita kertas berlubang	
 Magnetic Disk	IO yang menggunakan disk magnetik		 Manual Input	Input yang dimasukkan secara manual dan keyboard	
 Magnetic Drum	IO yang menggunakan drum magnetik		 Display	Output yang ditampilkan pada terminal	
			 Manual Operation	Operasi Manual	

SIMBOL	ARTI	CONTOH	SIMBOL	ARTI	CONTOH
 Keputusan	Keputusan dalam program		 Input / Output	Mengrepresentasikan Input data atau Output data yang diproses atau informasi.	
 Predefined Process	Rincian operasi berada di tempat lain		 Proses	Mengrepresentasikan operasi	
 Preparation	Pemberian harga awal		 Penghubung	Keluar ke atau masuk dari bagian lain flowchart khususnya halaman yang sama	
 Terminal Points	Awal / akhir flowchart		 Anak Panah	Mengrepresentasikan alir kerja	
 Punched card	Input / output yang menggunakan kartu berlubang		 Penjelasan	Digunakan untuk komentar tambahan	

Gambar 2.1 Simbol – simbol flowchart [34]

SIMBOL	ARTI	CONTOH
Communication Link 	Transmisi data melalui channel komunikasi, seperti telepon	
Off-line Storage 	Penyimpanan yang tidak dapat diakses oleh komputer secara langsung	

Gambar 2.2 Simbol – simbol *flowchart* (lanjutan) [34]

2.10. GOOGLE COLAB

Google Colab (Colaboratory) adalah sebuah layanan gratis dari *Google* yang memungkinkan pengguna untuk menulis dan mengeksekusi kode *Python* melalui *browser web*. Layanan ini berjalan di atas *platform cloud* milik *Google*, yang artinya tidak perlu menginstal atau mengkonfigurasi lingkungan *Python* di komputer lokal. *Google Colab* sangat populer di kalangan peneliti, pengembang, dan pembelajar karena menyediakan banyak fitur berguna, *Google Colab* adalah alat yang sangat berguna untuk pemula maupun profesional dalam bidang *data science, machine learning*, dan pengolahan data secara umum. Dengan akses ke sumber daya komputasi tinggi, kemudahan berbagi, dan lingkungan yang interaktif, *Google Colab* telah menjadi pilihan populer untuk banyak proyek dan tugas komputasi berbasis *Python* [35].

Google Colab adalah platform berbasis *cloud* yang memungkinkan pengguna menjalankan dan berbagi *notebook Jupyter* [36].

Colaboratory atau disingkat *Colab*, merupakan produk dari riset google. *Colab* memungkinkan siapa saja untuk menulis dan mengeksekusi kode *python* arbitrer melalui *browser*, dan sangat cocok untuk pembelajaran mesin, analisis data, dan pendidikan [37].

2.11. NOTEPAD

Notepad adalah salah satu perangkat lunak dari *windows* yang sudah mulai digunakan sejak *windows* versi 1.0 di tahun 1995. Aplikasi ini bisa dibilang simple dan jarang digunakan. Padahal, ada beberapa hal yang ternyata sangat terbantu sekali dengan adanya aplikasi *notepad* ini. *Notepad* adalah suatu aplikasi *default* dari *windows*. Fungsinya untuk membuat catatan kecil yang penting, atau mengedit teks dengan cepat [38].

Notepad / notepad ++ adalah sebuah penyunting teks dan penyunting kode sumber yang berjalan di sistem operasi *windows*. *Notepad* adalah salah satu program bawaan dari *windows* yang biasanya digunakan untuk menulis keterangan-keterangan yang penting dari program aplikasi seperti lisensi atau lainnya [39].

Notepad merupakan *IDE* yang paling sederhana dimiliki oleh bahasa system operasi *windows*. *IDE* ini memiliki kemampuan untuk membuat atau mengedit intaks Bahasa pemrograman, dengan kata lain *notepad* adalah *IDE* yang mempunyai kemampuan sebagai *text editor* .[40]

2.12. PENELITIAN SEJENIS

Tabel 2.2 Penelitian sejenis

1	Judul	Natural Language Processing untuk Analisis Sentimen Presiden Jokowi Menggunakan <i>Multi Layer Perceptron</i>
	Penulis	Nico Munasatya, Sendi Novianto [41]
	Masalah	Penelitian sebelumnya yang menggunakan metode serupa, penelitian ini mengatasi kelemahan pengumpulan data manual dengan menggunakan pendekatan yang lebih efisien.
	Solusi	Penelitian ini penting karena memanfaatkan <i>NLP</i> dan analisis sentimen untuk memahami opini publik di Twitter secara otomatis.
	Metode	<i>Multi Layer Perceptron</i>
	Hasil	Hasil Pengujian menunjukkan bahwa accuraci algoritma <i>Multi-layer perceptron</i> dengan jumlah <i>Hidden layer</i> , <i>Learning Rate</i> , <i>Shuffling</i> dan <i>Drop Out</i> berpengaruh dari nilai akurasi yang dihasilkan, Skenario 7 dengan menggunakan dataset presiden jokowi dan jumlah <i>Hidden layer</i> sebanyak 128 layer dan tidak menggunakan pengacakan data, sehingga mendapatkan perhitungan nilai akurasi terbaik pada akurasi atau Tes akurasi dengan persentase 93,26 % dan nilai <i>f1</i> atau <i>F-measure</i> dengan persentase 95,36 % juga nilai <i>Recall</i> terbaik dengan persentase 94.21 %.
2	Judul	Klasifikasi Sentimen Publik Terhadap Jenis Vaksin Covid-19 yang Tersertifikasi <i>WHO</i> Berbasis <i>NLP</i> dan <i>KNN</i>
	Penulis	Primandani Arsi, Iphang Prayoga, Muhammad Hasyim Asyari [42].
	Masalah	Bagaimana mengetahui sentimen publik di <i>Twitter</i> terhadap tiga jenis vaksin <i>Covid-19</i> yang telah mendapatkan lisensi dari <i>WHO</i> (<i>Pfizer</i> , <i>Moderna</i> , dan <i>AstraZeneca</i>) berdasarkan <i>NLP</i> dan algoritma <i>machine learning</i> .
	Solusi	Menggunakan metode <i>NLP</i> untuk <i>preprocessing</i> data teks, menentukan polaritas dan subjektivitas tweet, dan mengklasifikasikan sentimen menggunakan algoritma <i>KNN</i> .
	Metode	Melakukan ekstraksi data tweet dengan kata kunci <i>Pfizer</i> , <i>Moderna</i> , dan <i>AstraZeneca</i> , melakukan <i>preprocessing</i> dengan <i>NLP</i> , menghitung polaritas dan subjektivitas dengan <i>TextBlob</i> , dan melakukan klasifikasi dengan <i>KNN</i> berdasarkan skor polaritas.

	Hasil	Dari 3800 <i>tweet</i> yang dianalisis, ditemukan bahwa sentimen positif tertinggi ada pada <i>vaksin Pfizer</i> (47,30%), diikuti oleh <i>Moderna</i> (46,20%), dan terendah pada <i>AstraZeneca</i> (40,09%). Hal ini menunjukkan bahwa masyarakat lebih percaya pada <i>vaksin Pfizer</i> dibandingkan dengan <i>vaksin</i> lainnya.
3	Judul	Analisis Sentimen Opini Masyarakat Terhadap <i>Tech Winter</i> Pada <i>Twitter</i> Menggunakan <i>Natural Language Processing</i>
	Penulis	Rifky Aziz Fauzianto dan Supatman [43].
	Masalah	Fenomena <i>tech winter</i> yang menggambarkan keadaan dimana banyak startup teknologi mulai berguguran dan melakukan efisiensi, serta opini-opini masyarakat yang berkaitan dengan hal tersebut di <i>Twitter</i> .
	Solusi	Melakukan klasifikasi sentimen terhadap <i>tweet</i> yang mengandung kata ' <i>Tech Winter</i> ' dan ' <i>layoff</i> ' dalam Bahasa Indonesia, memanfaatkan kemampuan <i>Natural Language Processing (NLP)</i> .
	Metode	Menggunakan berbagai model <i>NLP</i> , yaitu <i>Regresi Logistik</i> , <i>Support Vector Machine (SVM)</i> , <i>Random Forest</i> , <i>Neural Network</i> , dan <i>Naive Bayes</i> , dan membandingkan efektivitas mereka dalam menafsirkan sentimen yang diungkapkan dalam <i>tweet</i> .
	Hasil	Model <i>SVM</i> dan <i>Naive Bayes</i> memberikan hasil yang sangat baik dalam hal akurasi dan metrik evaluasi lainnya, namun <i>Neural Network</i> menunjukkan keunggulan dalam validasi silang, yang menunjukkan konsistensi performanya yang tinggi dan kemungkinan kinerja yang lebih baik pada data yang belum dilihat.
4	Judul	Analisis Sentimen Kebijakan Penerapan Kurikulum Merdeka Sekolah Dasar dan Sekolah Menengah pada Media Sosial <i>Twitter</i> dengan Menggunakan Metode <i>Word Embedding</i> dan <i>Long Short-Term Memory Networks (LSTM)</i>
	Penulis	Alif Rizal Maulana, Satrio Hadi Wijoyo, dan Yusi Tyroni Mursityo [44].
	Masalah	Pandemi COVID-19 menyebabkan <i>learning loss</i> pada siswa dan mendorong Kemendikbudristek mengeluarkan kebijakan baru yaitu kurikulum merdeka. Kebijakan ini menimbulkan banyak opini dari berbagai pihak, terutama siswa yang terdampak langsung. Opini-opini ini sering disampaikan melalui media sosial <i>Twitter</i> , yang bisa dijadikan sumber data untuk analisis sentimen.
	Solusi	Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini siswa yang berkaitan dengan kurikulum merdeka di <i>Twitter</i> , dengan menggunakan pendekatan <i>machine learning</i> . Penelitian ini juga ingin menyajikan

		hasil analisis sentimen dalam bentuk dashboard visualisasi yang bisa digunakan oleh pihak sekolah dan dinas pendidikan sebagai bahan evaluasi dan refleksi.
	Metode	Data opini dimodelkan dengan arsitektur <i>deep learning Long Short-Term Memory Networks (LSTM)</i> dengan menggunakan <i>word embedding</i> sebagai teknik vektorisasi kata.
	Hasil	Penelitian ini berhasil mengumpulkan 455 data opini dari Twitter, yang terdiri dari 273 sentimen negatif, 101 sentimen netral, dan 81 sentimen positif. Model <i>LSTM</i> yang dibangun memiliki akurasi 81%, presisi 80%, recall 81%, dan f1-score 79%. <i>Dashboard</i> visualisasi yang dibuat menampilkan informasi mengenai persebaran data, kata terbanyak, opini terbanyak, dan hasil pengujian model. <i>Dashboard</i> ini mendapatkan <i>SUS Score</i> 62,5, yang menunjukkan tingkat <i>usability</i> yang cukup baik.
5	Judul	Implementasi Aplikasi Sentimen Pada Data <i>Twitter</i> Jelang Pemilu 2024
	Penulis	Choirul Humam dan Arif Dwi Laksito [45] .
	Masalah	Bagaimana menganalisis sentimen publik terhadap calon pemimpin dan isu-isu yang dibahas pada saat kampanye pemilu 2024 menggunakan data <i>Twitter</i> secara real-time.
	Solusi	Menggunakan metode <i>Deep Learning</i> algoritma <i>LSTM</i> untuk mengklasifikasikan <i>tweet</i> menjadi tiga kelas sentimen: mendukung, tidak mendukung, dan netral. Kemudian, mengimplementasikan aplikasi web yang dapat menampilkan grafik sentimen secara real-time.
	Metode	Menggunakan arsitektur <i>LSTM</i> untuk membangun model klasifikasi.
	Hasil	Model <i>LSTM</i> dengan metode <i>Random Oversampling</i> menghasilkan akurasi tertinggi sebesar 0.82 pada epoch 25. Aplikasi <i>web</i> dapat menampilkan grafik sentimen secara real-time dengan data <i>tweet</i> terbaru. Grafik sentimen menunjukkan perbandingan antara kelas mendukung, tidak mendukung, dan netral.

Berbagai penelitian yang telah dijabarkan menunjukkan bahwa penggunaan *Natural Language Processing (NLP)* dan teknik-teknik machine learning seperti *KNN (K-Nearest Neighbors)* serta *LSTM (Long Short-Term Memory Networks)* dapat memberikan kontribusi yang signifikan dalam analisis/klasifikasi sentimen terhadap berbagai konteks di media sosial, khususnya *Twitter*. Dalam berbagai konteks tersebut, berikut adalah beberapa kesimpulan umum:

a. **Efektivitas Metode:**

Penggunaan *multi-layer perceptron*, *KNN*, dan *LSTM* dalam berbagai penelitian menunjukkan efektivitasnya dalam menganalisis sentimen dengan tingkat akurasi yang tinggi.

b. **Penanganan Data:**

Penelitian cenderung mengatasi kelemahan pengumpulan data manual dengan memanfaatkan pendekatan yang lebih efisien, seperti pengumpulan data secara real-time melalui *web mining* atau *crawling tweet*.

c. **Penerapan dalam Konteks Tertentu:**

Penelitian-penelitian tersebut menunjukkan penerapan teknik klasifikasi sentimen dalam konteks yang beragam, seperti opini terhadap tokoh politik, vaksin, fenomena industri, kebijakan pendidikan, dan politik pemilu.

d. Pentingnya *NLP*:

Pentingnya penggunaan *NLP* sebagai pendekatan dalam *preprocessing* data teks untuk menghasilkan klasifikasi sentimen yang lebih akurat dan efisien.

e. Dampak Terhadap Keputusan dan Evaluasi:

Beberapa penelitian menunjukkan bahwa hasil klasifikasi sentimen dapat memberikan dampak pada pengambilan keputusan, evaluasi, dan refleksi terhadap kebijakan, opini masyarakat, dan isu-isu tertentu.