

**ANALISIS SENTIMEN PENGGUNA X TERHADAP PSSI
DAN TIMNAS INDONESIA DALAM PERKEMBANGAN
SEPAK BOLA DI INDONESIA DENGAN ALGORITMA
LSTM DAN BERT**

S K R I P S I



**AGUS PRASTIYO
F1E118028**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK ELEKTRO DAN INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS JAMBI
2025**

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini benar-benar karya saya sendiri. Sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang ditulis atau diterbitkan orang lain kecuali sebagai acuan atau kutipan dengan mengikuti tata penulisan karya ilmiah yang telah lazim.

Tanda tangan yang tertera dalam halaman pengesahan adalah asli. Jika tidak asli, saya siap menerima sanksi sesuai dengan peraturan yang berlaku.

Jambi, 18 Juli 2025
Yang menyatakan

AGUS PRASTIYO
F1E118028

**ANALISIS SENTIMEN PENGGUNA X TERHADAP PSSI
DAN TIMNAS INDONESIA DALAM PERKEMBANGAN
SEPAK BOLA DI INDONESIA DENGAN ALGORITMA
LSTM DAN BERT**

S K R I P S I

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana pada Program Studi Sistem Informasi



**AGUS PRASTIYO
F1E118028**

**PROGRAM STUDI SISTEM INFORMASI
JURUSAN TEKNIK ELEKTRO DAN INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS JAMBI
2025**

PENGESAHAN

Skripsi dengan Judul **“ANALISIS SENTIMEN PENGGUNA X TERHADAP PSSI DAN TIMNAS INDONESIA DALAM PERKEMBANGAN SEPAK BOLA DI INDONESIA DENGAN ALGORITMA LSTM DAN BERT”** yang disusun oleh **AGUS PRASTIYO, NIM : F1E118028**, telah dipertahankan di depan tim penguji pada tanggal 18 Juli 2025 dan dinyatakan lulus.

Susunan Tim penguji :

Ketua : Pradita Eko Prasetyo Utomo, S.Pd., M.Cs.
Sekertaris : Zainil Abidin, S.T., M.Eng.
Anggota : 1. Edi Saputra, S.T., M.Sc.
2. Reni Aryani, S.Kom., M.S.I.

Disetujui:

Pembimbing Utama

Pembimbing Pendamping

Pradita Eko Prasetyo Utomo, S.Pd., M.Cs.
NIP. 198710282019031010

Zainil Abidin, S.T., M.Eng.
NIP. 198903172022031006

Diketahui:

Dekan Fakultas Sains dan Teknologi

Ketua Jurusan Teknik Elektro
dan Informatika

Drs. Jefri Marzal, M.Sc., D.I.T.
NIP. 196806021993031004

Edi Saputra, S.T., M.Sc.
NIP. 198501082015041003

RINGKASAN

Sepak bola merupakan olahraga yang paling digemari di dunia, di Indonesia sepak bola juga termasuk olahraga yang paling di gemari, sehingga ketika ada hal yang berkaitan dengan sepak bola banyak orang yang memberikan opini terutama pada keputusan federasi dan performa Timnas ada setuju dan tidak mengenai kebijakan federasi, berpendapat dan mengkritik performa Timnas. Analisis sentimen digunakan untuk memahami opini dalam teks dan mengklasifikasikannya sebagai positif atau negatif. Teknik ini dapat membantu menggali opini publik dari media sosial, khususnya untuk melihat respon pengguna terhadap suatu topik. Dua metode yang sering digunakan dalam analisis sentimen adalah LSTM dan BERT. LSTM bekerja dengan mengingat informasi lampau untuk dijadikan acuan sebagai pengambilan keputusan. Namun tidak cukup baik dalam mengolah kata yang kompleks., sementara BERT merupakan model yang dibangun Google dari *transformers* agar bisa memahami maksud dari kata. Dari hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa akurasi yang dihasilkan oleh BERT lebih baik daripada LSTM dengan akurasi yang dihasilkan dari BERT yaitu sebesar 76%, sedangkan LSTM yaitu sebesar 71%.

RIWAYAT HIDUP



Agus Prastiyo lahir di Jambi, pada tanggal 29 Agustus 2000. Penulis merupakan anak tunggal dari pasangan Mujio dan Pairah. Jalur pendidikan formal yang ditempuh penulis adalah sebagai berikut:

1. SDN 149 Kota Jambi
2. SMPS Islam Al-Falah Kota Jambi
3. SMKS Revany Indra Putra Kota Jambi

Pada tahun 2018, peneliti melanjutkan pendidikan jenjang Strata 1 dan tercatat sebagai mahasiswa di Program Studi Sistem Informasi, Jurusan Teknik Elektro dan Informatika, Fakultas Sains dan Teknologi, Universitas Jambi melalui jalur SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri). Selama menempuh pendidikan S1, peneliti aktif dalam bidang akademik maupun non akademik. Pada tahun 2021, peneliti mengikuti program magang selama 3 bulan di Badan Pusat Statistik Kota Jambi. Dibawah bimbingan Bapak Pradita Eko Prasetyo Utomo, S.Pd., M.Cs. Sebagai pembimbing utama dan Bapak Zainil Abidin, S.T., M.Eng. sebagai Pembimbing Pendamping, peneliti menyelesaikan tugas akhir dengan judul “Analisis Sentimen Pengguna X Terhadap PSSI dan Timnas Indonesia Dalam Perkembangan Sepak Bola DI Indonesia Dengan Algoritma LSTM dan BERT”.

PRAKATA

Puji syukur kepada Allah SWT atas limpahan rahmat dan karunia-Nya peneliti dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Masyarakat Terhadap PSSI dan Timnas Indonesia Dalam Perkembangan Sepak Bola Indonesia” sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi. Dalam menyelesaikan skripsi ini peneliti tentunya tak lepas dari bantuan dan dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu dengan kerendahan hati dan rasa terima kasih, peneliti menyampaikan terima kasih kepada:

1. Kedua orang tua, Bapak Mujiyo dan Ibu Pairah yang mana adalah dua orang paling berharga dan motivasi terbesar bagi peneliti dalam menyelesaikan studi
2. Bapak Drs. Jefri Marzal, M.Sc., DIT. selaku Dekan Fakultas Sains dan Teknologi Universitas Jambi.
3. Bapak Edi Saputra, S.T., M.Sc.. selaku Ketua Jurusan Teknik Elektro dan Informatika Fakultas Sains dan Teknologi Universitas Jambi.
4. Ibu Reni Aryani, S.Kom., M.S.I. selaku Ketua Program Studi Sistem Informasi Universitas Jambi.
5. Bapak Pradita Eko Prasetyo Utomo, S.Pd., M.Cs. dan Bapak Zainil Abidin, S.T., M.Eng.. selaku Dosen Pembimbing yang telah banyak memberikan arahan dan dukungan dalam penyusunan skripsi.
6. Tim Penguji Skripsi, Bapak Edi Saputra, S.T., M.Sc.. Ibu Reni Aryani, S.Kom., M.S.I., yang telah memberikan berbagai masukan dan saran untuk kesempurnaan skripsi ini.
7. Bapak Zainil Abidin, S.T., M.Eng. selaku Dosen Pembimbing Akademik yang senantiasa memberikan motivasi dan pengarahan selama masa studi.
8. Seluruh Dosen Program Studi Sistem Informasi Universitas Jambi atas segala ilmu pengetahuan dan bimbingannya selama masa studi.
9. Kepada Yasril Ardiansyah yang telah banyak membantu memberi saran dan menemani selama menulis tugas akhir dan teman seperjuangan sistem informasi 2018 yang juga telah memberikan dukungan.
10. Dan semua orang yang seraca langsung maupun tidak langsung memberi dukungan yang tidak bisa di sebutkan satu persatu.

Semoga apa yang telah dilakukan dapat menjadi amal baik disisi Allah dan semoga Allah SWT membalas kebaikan yang telah diberikan kepada penulis. Selanjutnya, penulisan skripsi ini yang telah selesai dilakukan tidak

terlepas dari kekurangan. Untuk itu penulis terbuka untuk menerima saran dan masukan dari para pembaca untuk dapat memperbaiki tulisan ataupun untuk melanjutkan dan mengembangkan dari hasil penelitian ini. Semoga apa yang menjadi isi dari skripsi ini dapat menambah wawasan dan ilmu yang bermanfaat untuk kita semua.

Jambi 18 Juli 2025

Agus Prastiyo
F1E118028

DAFTAR ISI

PENGESAHAN	iii
RINGKASAN.....	iv
RIWAYAT HIDUP.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xiii
I. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian	5
II. TINJAUAN PUSTAKA	6
2.1 <i>Text Mining</i>	6
2.2 Media Sosial X	8
2.3 Sentimen Analisis	9
2.4 <i>Deep Learning</i>	9
2.5 <i>Natural Language Processing</i> (NLP)	10
2.6 <i>Long Short-Term Memory</i> (LSTM).....	11
2.7 <i>Bidirectional Encoder Representations from Transformers</i> (BERT)	13
2.8 Penelitian Terdahulu	14
III. METODOLOGI PENELITIAN	19
3.1 Waktu dan Tempat Penelitian	19
3.2 Alat dan Bahan Penelitian	19
3.3 Sumber Data.....	19
3.4 Kerangka Penelitian	19
3.5 Tahapan Pengumpulan Data	21
3.6 Tahapan Persiapan Penelitian	21
3.7 Metode Klasifikasi Data.....	25
3.8 Prediksi dan Evaluasi Model.....	26
3.9 Tahapan Visualisasi Data.....	27
IV. PEMBAHASAN.....	28
4.1 Pengumpulan Data	28
4.2 Pelabelan Data	30

4.3	Penanganan <i>Imbalance</i> Data	33
4.4	Preprocessing Data	34
4.5	<i>Encode</i> Label.....	36
4.6	<i>Split</i> Data	37
4.7	Tokenisasi.....	37
4.8	Penerapan Model Klasifikasi.....	41
4.9	<i>Training</i> Data	43
4.10	Implementasi Model Klasifikasi	45
4.11	Pengujian dan Evaluasi Model Klasifikasi.....	46
4.12	Visualisasi Data	47
4.13	Pembahasan	50
V.	KESIMPULAN DAN SARAN	51
5.1	Kesimpulan.....	51
5.2	Saran.....	51
	DAFTAR PUSTAKA.....	52
	LAMPIRAN	56

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu	15
Tabel 2. Hasil Terjemah	31
Tabel 3. Hasil Dari Pelabelan VADER	33
Tabel 4. Hasil Dari Membersihkan Data	36
Tabel 5. Hasil <i>Encode</i>	37
Tabel 6. Hasil Penerapan <i>One-Hot-Encoding</i>	39
Tabel 7. Hasil <i>Confision Matrix</i> LSTM	46
Tabel 8. Hasil <i>Confusion Matrix</i> BERT	46

DAFTAR GAMBAR

Gambar 1. Arsitektur <i>Deep Learning</i> (Sumber : Meints, 2019).....	10
Gambar 2. Struktur dari modul LSTM (Sumber : Maharani et al., 2024)	12
Gambar 3. Struktur BERT (Sumber : Devlin et al., 2019).....	14
Gambar 4. Kerangka Penelitian	20
Gambar 5. Contoh Hasil Penerapan Membersihkan Teks.....	23
Gambar 6. Perubahan Data <i>Encoding</i>	23
Gambar 7. Contoh Penerapan Tokenisasi LSTM	24
Gambar 8. Penerapan Tokenisasi BERT.....	24
Gambar 9. Kode Persiapan Proses <i>Crawling</i>	29
Gambar 10. Kode Program Melakukan <i>Harvest-Tweet</i>	30
Gambar 11. Hasil Pengumpulan Data dengan <i>Tweet-Harvest</i>	30
Gambar 12. Teks Bahasa Indonesia Ke Bahasa Inggris.....	31
Gambar 13. Kode Pelabelan dengan VADER	32
Gambar 14. Visualisasi Hasil Pelabelan dengan VADER	33
Gambar 15. Kode Program <i>Undersampling</i>	34
Gambar 16. Hasil dari <i>Random Undersampling</i>	34
Gambar 17. Kode Program Untuk Membersihkan Teks.....	36
Gambar 18. Kode Program <i>Encode</i>	36
Gambar 19. Kode Program Untuk <i>Split</i> Data.....	37
Gambar 20. Gambar Kode Program Tahap Tokenisasi	38
Gambar 21. Contoh Hasil Dari Tokenisasi pada LSTM.....	39
Gambar 22. Kode Program Penerapan <i>One-Hot-Encodig</i> :	39
Gambar 23. Kode Program Penerapan Tokenisasi BERT	40
Gambar 24. Contoh Hasil dari Penerapan Toenisasi BERT.....	40
Gambar 25. Kode Penerapan <i>Dataset Class Pytech</i>	41
Gambar 26. Kode Pemodelan LSTM	42
Gambar 27. Kode Program untuk Pemodelan BERT.....	43
Gambar 28. Kode Penerapan Training Model LSTM	43
Gambar 29. Kode Program <i>Training Args</i>	44
Gambar 30. Kode Printah <i>Trainer</i>	45
Gambar 31. Kode Perintah Penerapan Konfigurasi Pelatihan	45
Gambar 32. Kode Program Penerapan Fungsi pada Data Uji LSTM.....	45
Gambar 33. Kode Program Penerapan Fungsi pada Data Uji BERT.....	46
Gambar 34. Hasil Evaluasi Performa LSTM	47
Gambar 35. Hasil Evaluasi Performa BERT	47
Gambar 36. Kode Untuk Membuat <i>WordCloud</i>	48

Gambar 37. Hasil Visualisai Dengan <i>WordCloud</i>	48
Gambar 38. <i>WordCloud</i> Sentimen Negatif	49
Gambar 39. <i>WordCloud</i> Sentimen Positif	49

DAFTAR LAMPIRAN

Lampiran 1. <i>Source Code</i>	56
Lampiran 2. Hasil Terjemah Kata	72
Lampiran 3. Hasil Pelabelan Dengan VADER.....	87

I. PENDAHULUAN

1.1 Latar Belakang

Sepak bola merupakan olahraga yang paling populer di dunia karena diminati hingga mencapai 4 miliar orang penggemar di seluruh dunia (Saputra, 2025). Sepak bola merupakan permainan beregu di lapangan yang berbentuk persegi panjang yang menggunakan bola yang berukuran 68-70 cm dan dimainkan oleh dua kelompok yang saling berlawanan dan masing-masing kelompok terdiri atas sebelas orang pemain, dan permainan berlangsung selama 2 X 45 menit, kemenangan dari olahraga ini ditentukan oleh selisih gol yang masuk ke gawang lawan (*Arti Kata Sepak Bola - Kamus Besar Bahasa Indonesia (KBBI) Online*). Berdasarkan sejarah, awal mula sepak bola dikenal adalah pada abad pertengahan dan pada saat itu, permainan sepak bola diketahui mirip dengan permainan sepak bola rugby. Hingga pada tahun 1863 di London, Permainan sepak bola modern mulai diperkenalkan ketika klub sepak bola lokal di Inggris saat itu menerapkan berbagai aturan dalam permainan dan akhirnya terbentuklah asosiasi sepak bola. Beberapa aturan yang dibuat bertujuan agar klub sepak bola bisa bermain tanpa perselisihan yang membingungkan, dan aturan dibuat secara khusus untuk melarang memegang bola kecuali penjaga gawang. Dan hingga saat ini sepak bola telah menjadi salah satu cabang olahraga di olimpiade (*History of Football - The Origins*, 2012).

Sejarah sepak bola di Indonesia adalah pada masa Hindia Belanda, yaitu di Sumatra pada tanggal 16 November 1887 di Medan berdiri sebuah klub olahraga yang memainkan permainan sepak bola, tenis, dan atletik dengan nama Gymnastiek Vereniging, Tim ini tercatat sebagai klub sepak bola pertama yang lahir di Hindia Belanda pada awal tahun 1890, dan sempat bertanding dengan klub dari Penang dalam permainan kriket dan sepak bola. lalu pada 1 Juni 1899 berdiri sebuah klub sepak bola bernama Oostkust Sport Club Sumatra (OSCS) di Medan dan biasa disebut dengan Sport Club atau SOK. Dan pada tahun 1928 di Padang berdiri klub olahraga dengan nama Sport Vereniging Minang (SVM) yang bernaung organisasi sepak bola Padang yaitu Ilans Padang Electal (IPE), dan menjadi cikal bakal lahirnya PSP Padang.

Di Jawa sepak bola dimulai ketika didirikannya klub rekreasi dan sepak bola dengan nama Bataviasche Cricket-Football Club "Rood-Wit" pada tanggal 28 September 1893 di Batavia dan disetujui secara resmi pada Mei 1894, Di Surabaya juga berdiri sebuah klub dengan nama Victoria pada tahun 1894 dan pertandingan pertamanya dengan klub Sparta pada bulan Juni 1896 dengan hasil 6-1 untuk kemenangan Victoria. Dalam sistem kompetisi sepak bola

Hindia Belanda, hampir setiap kota memiliki sebuah federasi sepak bola. Federasi ini merupakan gabungan dari beberapa klub lokal dan kompetisi antar klub-klub lokal dijalankan oleh federasi tersebut. Turnamen sepak bola antar federasi kota pertama yang diadakan diwakili oleh federasi empat kota besar di Jawa, yaitu Batavia, Surabaya, Bandung, dan Semarang yang dilangsungkan pada akhir bulan Agustus tahun 1914 di Semarang dengan nama Koloniale Tentoonstellin (*Dutch East Indies - Football History*).

Saat ini, sepak bola di Indonesia berkembang dengan sangat baik, dimana sekarang Timnas Indonesia mampu berprestasi dengan berhasil lolos untuk menjalankan kualifikasi piala dunia putaran keempat, dan berdasarkan sejarahnya ini merupakan sebuah prestasi yang hebat, karena untuk pertama kalinya Indonesia bisa masuk ke putaran ketiga dan bisa berpeluang besar untuk bermain di piala Dunia yang akan dilaksanakan pada 2026. Piala Dunia sendiri merupakan puncak tertinggi dari kejuaraan sepak bola di seluruh dunia. Keberhasilan Indonesia saat ini tidak lepas dari peran banyak pihak, terutama PSSI sebagai organisasi yang bertanggung jawab dalam mengelola asosiasi sepak bola di Indonesia. PSSI merupakan asosiasi sepak bola Indonesia yang didirikan pada 19 April 1930 dengan Ketua Umum pertamanya adalah Ir. Soeratin Sosrosoegondo, dan pada 1952 PSSI resmi bergabung menjadi anggota dari badan sepak bola Dunia FIFA, kemudian PSSI juga bergabung dan menjadi anggota dari badan sepak bola Asia AFC, pada 1954. Selama perjalanannya dalam upaya mengembangkan sepak bola di Indonesia PSSI juga telah menggelar beragam kompetisi dan turnamen, seperti Liga Indonesia atau Liga 1, lalu Liga 2, Liga 3, Liga 1 Putri, Elite Pro Academy untuk jenjang usia muda U-16, dan U-18, dan Piala Soeratin (*Sejarah PSSI*).

PSSI masih sangat berperan dalam perkembangan sepak bola di Indonesia dengan keputusan dan kebijakan yang diambil dan dilakukan. Mulai dari peningkatan kualitas pada kompetisi yang ada di Indonesia dengan melakukan pembinaan terhadap wasit Liga 1. Selain peningkatan kompetisi di dalam negeri PSSI juga berperan untuk terus meningkatkan kualitas Timnas Indonesia. Karena itu, PSSI berwenang dalam menentukan staf dan pelatih yang akan akan bertugas untuk melatih Timnas selain itu, PSSI juga berperan untuk melakukan pemantauan terhadap pemain sepak bola yang memiliki potensi agar bisa terus berkembang dan bisa terus berprestasi, untuk meningkatkan kualitas dan prestasi Timnas Indonesia PSSI juga melakukan pemantauan terhadap para pesepak bola yang memiliki garis keturunan Indonesia yang mempunyai kemampuan, pengalaman dan potensi untuk membela Timnas Indonesia hingga bisa untuk di Naturalisi, hingga kebijakan dan keputusan

inilah yang diharapkan bisa berpengaruh pada performa Timnas Indonesia dalam pertandingan selanjutnya sehingga mendapatkan hasil yang maksimal.

Dengan keputusan dan kebijakan yang di ambil PSSI tersebut, berdampak terhadap munculnya harapan penggemar sepak bola di Indonesia agar Timnas Indonesia bisa tampil di Piala Dunia 2026. Meskipun Timnas Indonesia mendapatkan banyak pujian dan dukungan dari banyak pihak karena sudah menunjukkan prestasi dengan berhasil masuk kualifikasi Piala Dunia putaran keempat, Timnas Indonesia juga tidak luput dari kritikan dari berbagai pihak dan juga penggemar karena beberapa permainan yang tidak konsisten, mulai dari pemain, pelatih, hingga PSSI pun menjadi sasaran untuk di kritik dan yang terbaru PSSI pun mendapatkan banyak reaksi dari penggemar sepak bola di Indonesia karena melakukan pergantian pada pelatih untuk Timnas Indonesia, reaksi dari masyarakat itu di perlihatkan di berbagai media sosial diantaranya adalah Youtube, Tiktok, Instagram, Facebook, dan X,

Berdasarkan laporan *We Are Social dan Meltwater* ada 611,3 juta pengguna X di Dunia pada April 2024, dan Amerika Serikat merupakan Negara dengan pengguna X terbanyak dengan 106,23 juta pengguna X, Jepang menempati posisi kedua dengan 69,28 juta pengguna X, lalu posisi ketiga ditempati oleh India dengan 25,45 juta pengguna X, dan Indonesia menempati posisi keempat dengan pengguna X sebanyak 24,85 juta Pengguna (Stevany, 2024).

X merupakan layanan yang memungkinkan pengguna untuk saling berkomunikasi dengan pesan singkat sesuai dengan keinginan pengguna yang biasa di sebut sebagai *tweet*. *Tweet* dibatasi hingga 140 karakter panjangnya, dan telah diperluas hingga masih bisa berubah lagi di masa mendatang. Dengan begitu X merupakan layanan pesan teks global yang gratis dan berkecepatan tinggi (Russell & Klassen, 2019). Opini dan pikiran yang di *tweet* oleh pengguna bisa menjadi acuan dalam melihat reaksi dan juga sentimen mereka terhadap sebuah informasi atau bisa disebut analisis sentimen. Analisis sentimen merupakan istilah lain *opinion mining* yang merupakan salah satu cabang dari *text mining* yang bertujuan untuk menentukan persepsi dari publik terhadap suatu informasi dan juga bertugas mengklasifikasikan suatu teks ke dalam orientasi positif, negatif, dan netral. *Text mining* merupakan suatu proses analisis teks dari sebuah informasi yang penting untuk sebuah hasil yang spesifik (Azizah et al., 2022). Namun mesin hanya bisa memahami bahasa mesin seperti angka dan biner, mesin tidak dapat memahami bahasa manusia seperti teks, mesin masih bisa memahami bahasa manusia dengan *Natural Language Processing* (NLP), dengan algoritma ini mesin bisa memahami dan

memproses bahasa manusia, dan memungkinkan komputer untuk menganalisis bahasa manusia dan memungkinkan interaksi antara manusia dan mesin (Kulkarni & Shivananda, 2019).

Dalam menganalisis teks terdapat banyak banyak model diantaranya adalah LSTM dan BERT, LSTM merupakan arsitektur RNN yang telah dimodifikasi untuk menangani masalah *gradien* yang menghilang dan meledak serta untuk mengatasi masalah pelatihan data. LSTM dimodifikasi sedemikian rupa sehingga mencakup sel memori yang dapat menyimpan informasi dalam memori untuk jangka waktu yang lama. Dalam LSTM, serangkaian gerbang digunakan untuk mengontrol kapan informasi memasuki memori, yang memecahkan masalah *gradien* yang menghilang atau meledak (Manaswi, 2018). BERT merupakan pengembangan metode dari *transformer* yang di rilis pada tahun 2019, BERT hanya melakukan *encode* dan menghasilkan sebuah model bahasa, tidak seperti *transformer*, BERT hanya memerlukan *encoder* (Sofiana, 2024). Diantara keduanya LSTM merupakan modifikasi dari model yang sudah lama yaitu RNN, Sedangkan BERT merupakan pengembangan dari model yang terbilang cukup baru yaitu *transformer*, meskipun LSTM merupakan model lama tapi masih cukup relevan untuk tugas yang sederhana dan dataset yang sedikit, hal ini juga didukung dalam penelitian Ezen-cen (2020) “A Comparison of LSTM and BERT for Small Corpus” berdasarkan hasil penelitian untuk membangun *chatbot* dengan menggunakan dataset yang kecil, didapatkan hasil bahwa pemodelan LSTM memiliki nilai akurasi lebih baik dengan nilai akurasi 70,08%, dibandingkan pemodelan BERT yang memiliki nilai akurasi 67,15%.

Berdasarkan latar belakang diatas, peneliti akan menggunakan pemodelan LSTM dan BERT dalam melakukan klasifikasi sentimen analis data X, hal ini dilakukan untuk mengetahui perbandingan tingkat akurasi dalam melakukan klasifikasi sentimen analis dan juga mengetahui tanggapan masyarakat terhadap perkembangan sepak bola di Indonesia. Data yang di peroleh dari *tweet*, berdasarkan dua kata kunci “PSSI” dan “Timnas Indonesia” karna dua aspek tersebut bisa menjadi acuan tentang sejauh mana sepak bola Indonesia bisa berkembang. Data *tweet* akan diklasifikasikan menjadi tiga kategori sentimen yaitu positif, netral, dan negatif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, maka rumusan masalah yang akan diangkat dari penelitian ini adalah sebagai berikut:

1. Bagaimana perbandingan tingkat akurasi dari hasil klasifikasi sentimen terhadap perkembangan sepak bola di Indonesia menggunakan metode LSTM dan BERT?
2. Bagaimana hasil analisis sentimen dan implementasi metode *natural language processing* menggunakan metode LSTM dan BERT untuk melakukan klasifikasi sentimen analisis X?

1.3 Batasan Masalah

Agar penelitian tugas akhir ini tidak keluar dari pokok permasalahan maka ruang lingkup pembahasan dibatasi dengan masalah sebagai berikut :

1. Sumber opini yang digunakan pada wilayah Indonesia.
2. Analisis sentiment yang dilakukan menggunakan metode LSTM dan BERT menggunakan bahasa pemrograman Python dan tools *google colab*.
3. Analisis sentimen ini mencakup kelas positif, negatif dan netral.

1.4 Tujuan Penelitian

Tujuan dilakukan penelitian ini adalah :

1. Mengetahui perbandingan tingkat akurasi yang didapatkan dari analisis sentimen dengan metode LSTM dan BERT.
2. Mengetahui hasil dari analisis sentimen mengenai perkembangan sepak bola di Indonesia dengan metode LSTM dan BERT.

1.5 Manfaat Penelitian

Manfaat dilakukan penelitian ini adalah:

1. Memberikan pengetahuan mengenai perbandingan 2 metode klasifikasi data *tweets* opini pengguna media sosial X tentang perkembangan sepak bola di Indonesia menggunakan metode LSTM dan BERT.
2. Memberikan informasi sentimen masyarakat terhadap perkembangan sepak bola di Indonesia berupa sentimen positif, negatif dan netral.

II. TINJAUAN PUSTAKA

2.1 *Text Mining*

Dalam *teks mining* ada beberapa konsep dasar dan juga latar belakang yang perlu dipahami sebelum melakukan *teks mining* diantaranya adalah sebagai berikut:

Definisi *Text Mining*

Text mining didefinisikan sebagai proses mengekstraksi pengetahuan implisit dari data tekstual. Karena pengetahuan implisit yang merupakan hasil dari proses *text mining* tidak langsung tersedia dari data yang dikumpulkan, pengetahuan tersebut harus dibedakan dari informasi yang diambil. *Text mining* merupakan jenis khusus dari *data mining*, kita perlu mengeksplorasi *data mining* secara terkonsep, untuk memahaminya. *Data mining* merupakan proses mendapatkan pengetahuan implisit dari berbagai jenis data dalam konsep yang luas. Namun, dalam *data mining*, jenis data yang berperan sebagai sumber informasi dibatasi pada data relasional (George, 2022).

Keluaran dari *data mining* adalah pengetahuan implisit yang diperlukan secara langsung dalam pengambilan keputusan, dimana pengambilan beberapa variabel data yang relevan dengan kueri yang diberikan. Misalnya, dalam topik harga saham, prediksi harga saham di masa mendatang merupakan tugas umum *data mining*, sedangkan pengambilan sebagian harga saham masa lalu dan saat ini merupakan pengambilan informasi. Namun perlu diperhatikan bahwa kepastian yang sempurna tidak pernah ada dalam proses *data mining*. Perhitungan yang lebih canggih dan terbaru untuk membuat pengetahuan dari data mentah, yang disebut sintesis, diperlukan untuk melakukan tugas *data mining*.

Teks

Karena *text mining* berfokus dalam mengelola data berupa teks maka perlu untuk memahami tentang konsep teks. Bagian ini dibagi menjadi dua bagian :

1. Komponen Teks

Dalam bagian ini, teks didefinisikan sebagai kelompok kalimat atau paragraf yang ditulis dan disusun. Kata-kata individual disebut sebagai unit teks dasar, kata-kata tersebut digabungkan menjadi kalimat, secara tata bahasa, dan kalimat-kalimat tersebut disusun menjadi paragraf, secara logis. Di sini, kata dianggap sebagai unit teks dasar. Karena satu kata terdiri dari beberapa karakter, satu karakter juga dapat dianggap sebagai unit dasar.

Alasan penetapan kata sebagai unit dasar adalah karena karakter tidak memiliki makna dengan sendirinya, tetapi kata memiliki maknanya sendiri. Kata-kata yang digabungkan menjadi kalimat dengan aturan yang disebut tata bahasa, dan paragraf didefinisikan sebagai serangkaian kalimat yang diurutkan, kalimat-kalimat digabungkan menjadi paragraf berdasarkan urutan logis atau aturan semantik, bukan tata bahasa.

Sebuah teks dipandang sebagai sebuah artikel seperti artikel berita, artikel jurnal, paten, dan surat, jika teks tersebut terdiri dari satu paragraf, satu kelompok paragraf, dan beberapa kelompok paragraf. Abstrak dari sebuah artikel teknis atau artikel berita dengan teksnya sendiri bisa disebut sebagai teks pendek. Sebuah teks yang memiliki satu kelompok paragraf yang disebut bagian dapat dianggap sebagai teks sedang. Sebuah teks dengan beberapa kelompok paragraf sebagai artikel multi-bagian dianggap sebagai teks panjang.

2. Format Teks

Berbagai format teks tersedia dalam *teks mining*. Format teks yang populer adalah yang dibuat oleh perangkat lunak MS Office, seperti file MS Word yang ekstensinya adalah “*doc*,” file MS PowerPoint yang ekstensinya adalah “*ppt*,” dan file MS Excel yang ekstensinya adalah “*xls*.” Teks cenderung paling populer ditransfer dalam format PDF melalui Internet antara pengguna dengan mengunggah dan mengunduhnya. Teks biasa yang disebutkan sebagai format paling sederhana adalah teks yang dibuat oleh editor teks. Setiap teks biasanya sesuai dengan filenya sendiri yang ekstensinya adalah “*txt*”. Pengumpulan data biasa ditetapkan sebagai proses *text mining* dalam penerapan sistem mengkategorikan dan pengelompokan teks (Jo, 2024).

Pengindeksan Teks

Pengindeksan teks didefinisikan sebagai proses mengubah teks-teks menjadi daftar kata. Karena teks atau teks-teks diberikan sebagai bentuk-bentuk yang tidak terstruktur dengan sendirinya atau pada dasarnya, hampir tidak mungkin untuk memproses bentuk mentahnya secara langsung dengan menggunakan program komputer. Dengan kata lain, pengindeksan teks berarti proses segmentasi teks yang terdiri dari kalimat-kalimat menjadi kata-kata yang disertakan. Daftar kata adalah hasil dari pengindeksan teks sebagai keluaran dari pengindeksan teks, dan akan menjadi masukan untuk representasi teks. Teks pada dasarnya adalah data tidak terstruktur yang tidak seperti data numerik, sehingga program komputer tidak dapat memprosesnya dalam bentuk mentahnya. Tidak mungkin untuk menerapkan operasi numerik pada teks dan tidak mudah untuk mengodekan teks menjadi nilai numeriknya sendiri. Suatu teks merupakan untaian yang terlalu panjang (Jo, 2024).

Terdapat beberapa langkah dalam melakukan pengindeksan text diantaranya adalah:

1. Tokenisasi

Tokenisasi didefinisikan sebagai proses segmentasi teks atau teks-teks menjadi token berdasarkan spasi atau tanda baca.

2. *Stemming*

Stemming merupakan proses mengubah kata menjadi kata dasar tanpa ada imbuhan.

3. Penghapusan kata henti/ sambung

Merupakan proses menghapus kata yang tidak perlu seperti “dan” dan “yang”.

Selain dari 3 langkah diatas disebutkan terdapat beberapa langkah lain yang bisa ditambahkan dan di sesuaikan dengan teks dan tujuan dari penelitian.

2.2 Media Sosial X

X adalah layanan yang memungkinkan pengguna saling berkomunikasi dengan mengirim pesan singkat yang sesuai dengan isi pikiran atau ide dari pengguna. Secara historis, *tweet* ini dibatasi hanya sampai 140 karakter panjangnya dalam sekali kirim, namun batasan ini masih bisa diperluas dan bisa saja berubah lagi di masa yang akan. Dalam hal itu, X bisa didefinisikan sebagai layanan pesan teks global yang gratis dan memiliki kecepatan yang tinggi. Dengan kata lain, ini adalah bagian dari infrastruktur yang memungkinkan komunikasi yang sangat cepat, mudah dan juga memungkinkan pengguna untuk berbagi opini singkat dengan cepat tanpa harus saling mengikuti dengan pengguna lain. Model pengikut asimetris inilah yang menjadikan X lebih seperti grafik minat daripada jejaring sosial lain. Sementara beberapa situs web sosial seperti Facebook dan LinkedIn mengharuskan penerimaan bersama atas hubungan antara pengguna, model hubungan X memungkinkan untuk mengikuti perkembangan terbaru dari pengguna lain, meskipun pengguna lain tersebut mungkin tidak memilih untuk saling mengikuti. Model pengikut X sederhana tetapi mengeksplorasi aspek mendasar dari apa yang membuat kita menjadi manusia rasa ingin tahu kita. Baik itu kegilaan terhadap gosip dari selebriti favorit, keinginan untuk mengikuti tim olahraga favorit, minat yang besar terhadap topik politik tertentu, atau keinginan untuk terhubung dengan seseorang yang baru, X memberi peluang tak terbatas untuk memuaskan rasa ingin tau pengguna (Russell & Klassen, 2019).

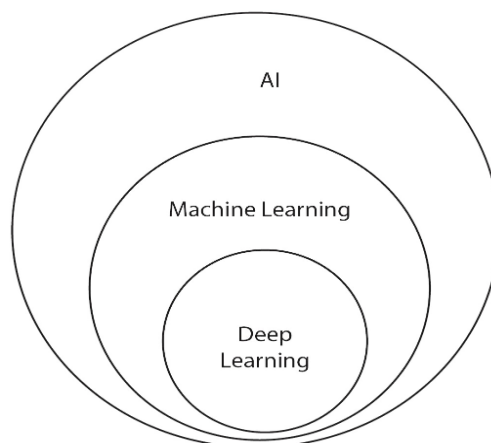
2.3 Sentimen Analisis

Sentimen Analisis merupakan proses klasifikasi opini menjadi beberapa sikap, positif, netral, atau negatif. Opini diberi data masukan tekstual, dan salah satu dari tiga sikap dihasilkan sebagai keluaran. Sentimen Analisis digunakan untuk mengklasifikasikan opini secara otomatis terhadap sebuah produk komersial atau isu politik berdasarkan sikap. Analisis ini perlu dibedakan dari penentuan topik, di mana teks diklasifikasikan berdasarkan satu atau beberapa topik. Ada tiga kategori dalam analisis sentimen. Positif diartikan sebagai opini yang menggambarkan sesuatu atau seseorang dengan ekspresi yang positif atau baik, luar biasa, dan hebat. Netral berarti menggambarkan sesuatu secara objektif tanpa positif maupun negatif, atau dengan campuran keduanya. Negatif berarti menggambarkan sesuatu dengan ekspresi negatif atau buruk, benci, tidak tertarik, dan tidak suka (Liu, 2012).

Netral dapat dibagi menjadi dua jenis: satu tanpa ekspresi sentimen dan satu dengan campuran positif dan negatif. Algoritma *machine learning* dapat diterapkan pada analisis sentimen dengan menerapkan tugas klasifikasi. Teks yang diberi label dengan salah satu dari tiga kategori dikumpulkan dan dikodekan ke dalam vektor numerik, sehingga ketika teks baru dimasukkan maka akan di klasifikasi berdasarkan algoritma *machine learning*. Analisis sentimen cenderung sangat bergantung pada istilah positif dan negatif. Karena pendapat diberikan sebagai teks yang sangat pendek, informasi tidak cukup untuk membedakan pendapat positif dan negatif satu sama lain. Suatu opini atau topik cenderung mengandung ungkapan sehari-hari dan bahasa gaul. Opini negatif yang diungkapkan dengan lembut dalam kata-kata netral dapat salah diklasifikasikan ke dalam opini netral (Pang & Lee, 2008).

2.4 Deep Learning

Deep Learning merupakan salah satu cabang dari *machine learning* yang menekankan pada penggunaan jaringan saraf tiruan (*artificial neural networks*) dengan banyak lapisan (*deep neural networks*). Pendekatan ini menjadi sangat populer karena kemampuannya dalam memproses data dalam jumlah besar dan menyelesaikan tugas-tugas kompleks seperti pengenalan gambar, pemrosesan bahasa alami, dan pengenalan suara. *Deep learning* menggunakan arsitektur yang terdiri dari beberapa lapisan tersembunyi (*hidden layers*) untuk mengekstraksi fitur secara bertingkat dari data mentah. Tiap lapisan memiliki kemampuan untuk mempelajari representasi data yang lebih abstrak, yang memungkinkan model mengenali pola-pola kompleks (Yoshua Bengio et al., 2015).



Gambar 1. Arsitektur *Deep Learning* (Sumber : Meints, 2019)

Salah satu keunggulan utama dari *deep learning* adalah kemampuannya dalam pembelajaran fitur secara otomatis (*automatic feature learning*). Jika pada metode *machine learning* tradisional diperlukan proses ekstraksi fitur secara manual, maka pada *deep learning*, proses ini dilakukan secara otomatis oleh jaringan. *Deep learning* sangat efisien dalam mengenali pola yang sangat kompleks dalam data yang tidak terstruktur (Schmidhuber, 2015).

2.5 *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada interaksi antara komputer dan bahasa manusia secara alami. Perkembangan NLP mengalami kemajuan yang sangat pesat, terutama berkat meningkatnya kemampuan komputasi, ketersediaan data besar (*big data*), dan juga terintegrasi dengan teknologi *deep learning*. NLP tidak hanya digunakan dalam penerjemahan bahasa otomatis, tetapi juga dalam *chatbot*, analisis sentimen, sistem pencarian informasi, hingga pengenalan suara (Young et al., 2018).

NLP sangat bergantung pada pendekatan statistik dan model berbasis jaringan saraf. Hal ini berbeda dengan pendekatan lama yang lebih banyak menggunakan aturan linguistik buatan manusia (*rule-based*). Kemajuan ini dipercepat dengan hadirnya algoritma seperti *word embeddings* (contohnya *Word2Vec* dan *GloVe*), yang memungkinkan komputer memahami konteks kata dalam bentuk vektor angka (Jurafsky & Martin, 2021).

NLP mulai beralih ke model berbasis *deep learning*. Salah satu tonggak penting adalah pengenalan model *Recurrent Neural Network (RNN)* dan *Long Short-Term Memory (LSTM)* yang dapat menangani data sekuensial seperti teks dan suara. Namun, model ini memiliki keterbatasan dalam pemrosesan konteks panjang. Untuk mengatasi hal ini, para peneliti mengembangkan model

transformer yang menjadi dasar bagi model NLP besar seperti BERT dan GPT (Yoshua Bengio et al., 2015).

Model-model ini menggunakan mekanisme *self-attention* yang memungkinkan mereka untuk memahami hubungan antar kata dalam satu kalimat maupun antar kalimat secara lebih efisien (Vaswani et al., 2017). Selain itu, pelatihan model menggunakan teknik *transfer learning* juga menjadi populer, yaitu dengan melatih model pada data besar terlebih dahulu, kemudian disesuaikan (*fine-tuned*) untuk tugas tertentu (Howard & Ruder, 2018).

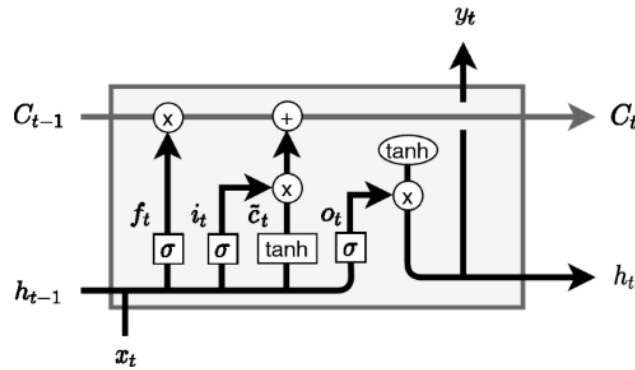
NLP menunjukkan kemajuan dalam pengolahan Bahasa non-Inggris, termasuk Bahasa Indonesia. Dengan adanya corpus digital yang semakin besar dan model multibahasa seperti *mBERT* dan *IndoBERT*, kualitas NLP dalam bahasa lokal semakin meningkat. Dengan demikian, perkembangan NLP sangat dipengaruhi oleh kemajuan *deep learning* dan model neural networks modern. Integrasi antara NLP dan teknologi terkini telah memungkinkan sistem komputer memahami dan merespons bahasa manusia secara lebih cerdas dan alami (Cahyawijaya et al., 2021).

2.6 Long Short-Term Memory (LSTM)

(LSTM) merupakan salah satu jenis *deep learning* yang dirancang untuk mengatasi masalah *vanishing gradient* yang umum terjadi pada LSTM standar saat memproses data berurutan yang panjang. *Recurrent Neural Network* (LSTM) merupakan sebuah arsitektur *artificial neural network* yang dirancang secara khusus untuk memproses data yang berurutan, seperti teks dan sinyal suara. Keunggulan utama dari LSTM adalah kemampuannya untuk mempertahankan informasi dari langkah-langkah sebelumnya dalam memori internal, sehingga sangat cocok untuk aplikasi dalam NLP (Yoshua Bengio et al., 2015). Namun, LSTM memiliki kelemahan dalam mengatasi *long-term dependencies* karena adanya masalah *vanishing gradient*. Hal ini menyebabkan LSTM kurang efektif dalam memproses data berurutan yang panjang (Yoshua Bengio et al., 2015).

LSTM muncul dengan kemampuan untuk mengingat informasi penting yang terdapat pada data *input* selama beberapa waktu tertentu, sehingga sangat efektif untuk berbagai tujuan seperti prediksi deret waktu NLP. Hal ini bisa terjadi karena LSTM memiliki struktur khusus berupa sel memori yang bisa mempertahankan keadaan selama beberapa waktu dan beberapa gerbang (*gates*) yang berfungsi untuk mengatur aliran informasi yang masuk dan keluar dari sel tersebut (Hochreiter & Schmidhuber, 1997). Gerbang-gerbang ini terdiri dari *input gate*, *forget gate*, dan *output gate* yang berfungsi sebagai penentu informasi mana yang perlu disimpan, dibuang, atau dikeluarkan.

LSTM memiliki struktur seperti rantai, tetapi modul berulangnya memiliki struktur yang berbeda. Alih-alih memiliki satu lapisan jaringan saraf, terdapat empat lapisan yang berinteraksi dengan cara yang sangat khusus (*Understanding LSTM Networks -- Colah's Blog*). Berikut adalah arsitektur LSTM yang terdiri dari beberapa komponen utama, yaitu sebagai berikut:



Gambar 2. Struktur dari modul LSTM (Sumber : Maharani et al., 2024)

1. *Cell state*, berfungsi sebagai jalur utama untuk informasi yang mengalir secara terus-menerus.
2. *Forget gate*, berfungsi sebagai penentu informasi mana yang harus dilupakan dari cell state.
3. *Input gate*, berfungsi untuk mengontrol informasi baru yang akan ditambahkan ke cell state.
4. *Output gate*, berfungsi untuk menentukan informasi apa yang akan diteruskan ke langkah selanjutnya atau ke output jaringan.

Melalui struktur diatas, LSTM dapat mempertahankan konteks data selama proses pelatihan, yang sangat berguna dalam tugas yang membutuhkan pemahaman konteks jangka panjang, seperti NLP dan prediksi (Graves, 2013). LSTM banyak digunakan dalam aplikasi yang melibatkan data sekuensial. Dalam bidang *Natural Language Processing* (NLP), LSTM efektif digunakan untuk analisis sentimen, terjemahan, dan juga pengenalan suara karena kemampuannya dalam memahami konteks kata dalam kalimat (Sundermeyer et al., 2012). Di bidang keuangan, LSTM bisa digunakan untuk prediksi harga saham dan analisis pasar dengan menggunakan data historis secara akurat.

Namun, LSTM juga memiliki kelemahan, seperti kompleksitas komputasi yang tinggi dan waktu pelatihan yang lebih lama. Oleh karena itu, dalam beberapa kasus, arsitektur yang lebih baru seperti *transformer* mulai menggantikan LSTM karena efisiensi dan akurasi yang lebih baik (Vaswani et al., 2017).

2.7 ***Bidirectional Encoder Representations from Transformers (BERT)***

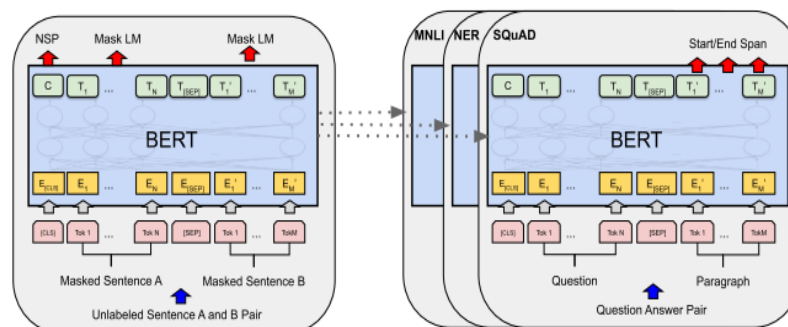
BERT (*Bidirectional Encoder Representations from Transformers*) merupakan sebuah model *deep learning* yang berbasis *transformer* yang diperkenalkan oleh Google pada tahun 2018. Model ini dirancang untuk memahami konteks kata dalam sebuah kalimat dengan memperhatikan konteks dari dua arah, yaitu kiri dan kanan, sehingga bisa menangkap makna kata lebih efektif dibandingkan model sebelumnya yang bersifat satu arah. Dalam konteks *Natural Language Processing*, BERT telah merubah cara mesin memahami teks dengan memberikan representasi kata yang lebih baik dan dinamis. BERT menggunakan dua metode yaitu *pre-training* dan *fine-tuning*. Pada tahap *pre-training*, model belajar dari sejumlah besar data teks tanpa label dengan dua tugas utama: *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). MLM bertujuan memprediksi kata yang disembunyikan dalam sebuah kalimat, sedangkan NSP bertugas menentukan apakah dua kalimat berurutan secara logis. Pendekatan ini membuat BERT mampu memahami konteks kalimat secara mendalam dan menghubungkan makna antar kalimat. BERT mampu memproses konteks dua arah sekaligus sehingga menghasilkan representasi yang lebih akurat untuk tugas-tugas NLP seperti klasifikasi teks, ekstraksi informasi, dan penjawaban pertanyaan (Q&A) (Devlin et al., 2019).

Salah satu komponen utama yang membuat BERT sangat kuat adalah arsitektur berbasis *transformer* yang terdiri dari beberapa *layer encoder*. *Layer-layer* ini bertugas memproses data *input* secara bertahap dan memungkinkan model memahami konteks secara mendalam. BERT terdiri dari beberapa lapisan *encoder transformer* yang disusun secara bertumpuk. Setiap *layer* memiliki fungsi utama untuk melakukan *self-attention* dan *feed-forward neural network*. *Layer-layer* ini secara berturut-turut memproses token *input* dengan cara memperhatikan hubungan antar token baik dari kiri maupun kanan, sehingga menghasilkan representasi yang kaya konteks. menjelaskan bahwa BERT menggunakan 12 atau 24 *layer encoder* tergantung dari versi yang digunakan (BERT-base atau BERT-large). Setiap *layer* dilengkapi dengan mekanisme residual connection dan layer normalization untuk menjaga kestabilan pelatihan (Devlin et al., 2019).

Setiap *layer* pada BERT bertanggung jawab untuk mengekstrak fitur dari teks *input* secara berlapis-lapis. Layer awal fokus pada pemahaman kata secara lokal, sementara layer lebih dalam menangkap makna yang lebih abstrak dan kontekstual. Hal ini memungkinkan BERT untuk memahami nuansa bahasa, seperti *polisemi* (kata dengan makna ganda) dan struktur kalimat kompleks. Sebagai contoh, *layer self-attention* mengkalkulasi bobot perhatian (*attention*

weights) untuk setiap token terhadap token lain dalam kalimat, memungkinkan model untuk memahami hubungan penting antar kata. Setelah itu, *feed-forward network* memproses informasi tersebut untuk memperkuat pola yang relevan (Tenney et al., 2020).

Layer pada metode BERT merupakan pondasi utama yang memungkinkan model memahami konteks kalimat secara mendalam dan fleksibel melalui proses self-attention dan transformasi non-linear. Struktur bertingkat ini menjadi kunci utama keberhasilan BERT dalam berbagai tugas NLP seperti klasifikasi teks, penerjemahan, dan ekstraksi informasi. Secara umum, BERT terdiri dari *encoder transformer* bertumpuk (*stacked transformer encoder*). Setiap lapisan *encoder* memiliki *self-attention* dan *feedforward layer*. *Input* ke BERT terdiri dari tokenisasi khusus menggunakan WordPiece dan ditambahkan dengan token spesial seperti [CLS] dan [SEP]. Model ini dilatih menggunakan corpus besar seperti Wikipedia dan BookCorpus untuk menghasilkan representasi kata yang kontekstual. BERT tidak menghasilkan teks, tetapi digunakan untuk mendapatkan representasi yang kaya dari teks yang dimasukkan untuk kemudian digunakan dalam berbagai tugas *downstream* NLP (Devlin et al., 2019). Berikut Gamabar struktur BERT:



Gambar 3. Struktur BERT (Sumber : Devlin et al., 2019)

2.8 Penelitian Terdahulu

Dalam melakukan penelitian, penulis membutuhkan acuan dari penelitian terdahulu yang berkaitan dengan penelitian, tujuannya adalah untuk mempermudah dan menjadi pembanding dari penelitian yang penulis lakukan. Berikut merupakan penelitian yang berkaitan dengan penelitian yang akan dilakukan oleh penulis.

Tabel 1. Penelitian Terdahulu

No	Peneliti/Tahun	Topik	Keeterangan
1	Azizah, et al. (2022)	Analisis Sentimen Masyarakat terhadap Kebijakan Vaksinasi Covid-19 di Indonesia pada X Menggunakan Algoritma LSTM	menganalisis sentimen masyarakat mengenai kebijakan vaksinasi di Indonesia dengan menggunakan algoritma LSTM. Penelitian ini menggunakan data yang berasal dari X Berdasarkan hasil penelitian, dapat ditarik kesimpulan bahwa sentimen masyarakat mengenai 'kebijakan vaksinasi di Indonesia' cenderung lebih banyak sentimen positif daripada sentimen negatif dan sentiment netral dengan menghasilkan menghasilkan 79% <i>tweet</i> positif, 13% <i>tweet</i> netral dan 8% <i>tweet</i> negatif.
2	Farida & Rochmawati, (2024)	Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di X	penelitian ini dilakukan dengan tujuan untuk mengetahui analisis sentimen terhadap fenomena menggunakan metode <i>Long Short Term Memory</i> dengan <i>Bidirectional Encoder Representations from Transformers</i> di X, ini menjadi kombinasi model yang dapat meningkatkan akurasi. Data yang digunakan yaitu berjumlah 3000 data dari X. Pada pengujian model dilakukan dengan pengujian beberapa parameter untuk mendapatkan model yang optimal diantaranya ukuran batch size

			128, dropout 0.5, dense layer 32 dan LSTM layer 64. Hyperparameter tersebut dilakukan pelatihan model yang menghasilkan performa terbaik dengan akurasi sebesar 0.9585, f1-score 0.9589, loss 0.1001, kemudian model dapat memprediksi tweets childfree dan menghasilkan precision sebesar 0.7839, recall 0.77, dan f1-score 0.7697.
3	Khadapi & Maruli Pakpahan, (2024)	Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi X tentang Pemilu 2024	peneliti mengidentifikasi tujuan analisis sentimen yang ingin dicapai, seperti memahami pola sentimen masyarakat terkait Pemilu 2024, peneliti mengumpulkan data dari platform X dengan menggunakan kata kunci yang relevan untuk diskusi seputar Pemilu 2024. Peneliti juga melakukan analisis sentimen awal untuk mendapatkan gambaran umum mengenai sentimen yang berkembang di kalangan pengguna X. Hasil dari penelitian secara keseluruhan, penelitian ini menegaskan bahwa baik LSTM maupun BERT memiliki potensi yang signifikan dalam analisis sentimen, tetapi setiap model perlu disesuaikan dengan karakteristik data dan tujuan spesifik dari aplikasi yang diinginkan. Melalui pengembangan lebih lanjut dan

			integrasi teknik yang sesuai, kedua model ini dapat terus berkontribusi dalam peningkatan akurasi klasifikasi sentimen serta aplikasi dalam berbagai bidang lain yang memerlukan pemrosesan bahasa alami.
4	Maharani, et al., (2024)	Analisis sentimen vaksin covid-19 pada X menggunakan <i>recurrent neural network</i> (LSTM) dengan algoritma <i>long short-term memory</i> (LSTM)	penelitian ini akan menggunakan <i>Recurrent Neural Network</i> dengan algoritma <i>Long Short-Term Memory</i> untuk analisis sentimen terkait vaksin Covid-19. Pengambilan data pada penelitian ini dilakukan dengan X <i>Crawling</i> . hasil bahwa klasifikasi menggunakan <i>Recurrent Neural Network</i> dengan algoritma <i>Long Short-Term Memory</i> menghasilkan akurasi sebesar 80,34%. Masyarakat Indonesia cenderung beropini positif terhadap vaksin Covid-19
5	Imawan,et al. (2025)	Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert	Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap rencana kenaikan PPN sebesar 12% dengan menggunakan model BERT. Hasil penelitian menunjukkan dominasi sentimen negatif terhadap rencana kenaikan PPN 12%, yang mencerminkan kekhawatiran masyarakat terhadap kenaikan biaya hidup dan penurunan daya beli. Model BERT terbukti efektif

dalam analisis sentimen berbasis teks media sosial dengan *accuracy* 83% serta nilai *precision*, *recall*, dan *F1-score* rata-rata masing-masing 83%, 82%, dan 82%.

Sumber : Penelitian Terdahulu

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Tempat penelitian ini dilakukan adalah di Kota Jambi. Dan data yang diperoleh merupakan hasil *crawling* dari data X melalui *tweet-harvest* dari 27 juni 2024 sampai 1 mei 2025.

3.2 Alat dan Bahan Penelitian

Peralatan dan bahan yang digunakan dalam menunjang proses penelitian

agar dapat mudah dilakukan adalah sebagai berikut :

1. Perangkat Keras (*Hardware*)

- a. Laptop acer
- b. *Processor* Intel (R) Core (TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
- c. HDD 1 TB
- d. RAM 12 GB
- e. *system type* 64-bit *operating system*

2. Perangkat Lunak (*Software*)

- a. Sistem operasi *Windows 10*.
- b. Bahasa pemrograman Python 3
- c. *Google colaboratory*
- d. *Library regex, pandas, NLTK, matplotlib, keras, transformers, train_test_split, confusion_matrix, googletrans, sklearn, tensorflow.*

Penelitian ini membutuhkan data opini dari pengguna X di Indonesia yang didapatkan dari sosial media X, Pengetahuan lain juga didapatkan melalui bahan bacaan yang sesuai dengan topik penelitian seperti buku, jurnal dan literatur diinternet.

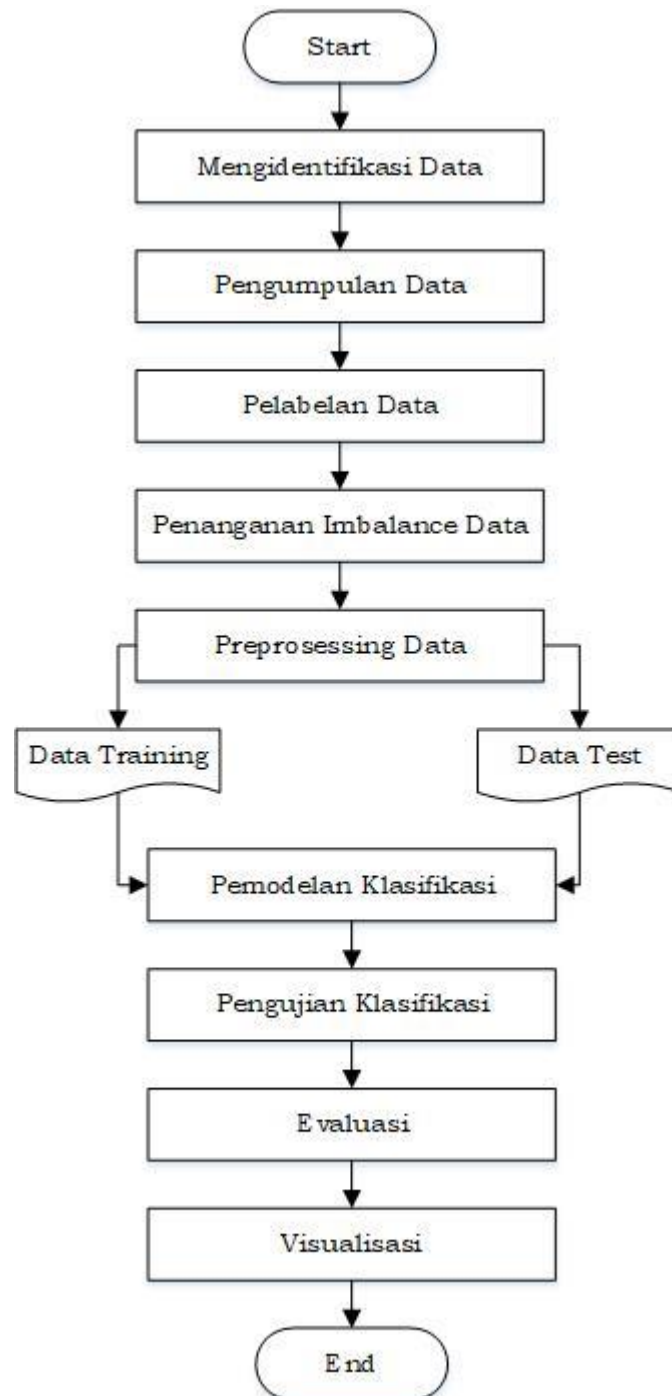
3.3 Sumber Data

Data yang digunakan dalam penelitian ini bersumber dari opini masyarakat di media sosial X kepada PSSI dan Timnas Indonesia terhadap perkembangan sepak bola di Inonesia. Pengumpulan data dilakukan dengan API X (*tweet-harvest*) untuk mengambil data *tweet* dari pengguna X di Indonesia dengan kata kunci “PSSI” dan “Timnas Indonesia”.

3.4 Kerangka Penelitian

Kerangka penelitian berperan untuk menunjukkan alur kerja dari penelitian yang akan dilakukan. Berikut kerangka penelitian untuk menganalisa sentimen dari data *tweets* menggunakan metode LSTM dan BERT

yang dilakukan dalam beberapa tahapan, Berikut tahapan yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 4.



Gambar 4. Kerangka Penelitian

Dalam kerangka penelitian langkah pertama adalah mengidentifikasi masalah, dalam hal ini Indentifikasi masalah yang didapatkan adalah untuk mengklasifikasi opini masyarakat Indonesia di media sosial X mengenai PSSI dan Timnas Indonesia terkait perkembangan sepak bola di Indonesia dan

membandingkan akurasi antara LSTM dan BERT dalam melakukan klasifikasi data X, selanjutnya adalah pengumpulan data dengan cara *crawling* dari API *Tweet-Harvest*, setelah data di kumpulkan, lalu data diterjemahkan kedalam Bahasa Inggris, tahap selanjutnya adalah pelabelan data, proses ini menggunakan *tool* *Lexion-VADER* untuk memberi label pada data, setelah itu data yang sudah diberikan label terdapat ketidakseimbangan satu sama lain dan dilaakukan *random undersampling* untuk menyetarakan dengan data minoritas, kemudian data dilakukan *preprocessing* dengan melakukan tokenisasi dan membagi data menjadi data 80% data *training* dan 20% data *test*, setelah data dibagi tahap selanjutnya membangun model klasifikasi LSTM dan BERT dengan data *training*, kemudian model akan di *training* menggunakan data *test*. setelah itu hasil dari *training* diuji untuk mengetahui tingkat akurasi dan *confusion matrix*. dan akan didapatkan perbandingan akurasi antar model.

3.5 Tahapan Pengumpulan Data

Pengumpulan data merupakan tahap penting dalam proses penelitian sentimen analis, karena kualitas data sangat memengaruhi hasil yang sentiment yang diperoleh. Dalam penelitian yang berbasis data digital, yang memanfaatkan media sosial X, teknik pengumpulan data mengalami perkembangan yang pesat seiring kemajuan teknologi informasi. Oleh karena itu, penelitian kali ini menggunakan salah satu metode yaitu *tweet-harvest* karena mampu mengakses *tweet* publik secara lebih fleksibel. Penelitian ini mengambil *tweet* pengguna X di Indonesia dengan kata kunci “PSSI” dan “Timnas Indonesia”. Hasil dari *tweet-harvest* terdapat beberapa struktur data diantaranya *tweet_id*, *text*, *username*, *created_at*, *retweet_count*, *like_count*, dan *lang*.

3.6 Tahapan Persiapan Penelitian

Sebelum membangun dan *training* model klasifikasi maka terdapat beberapa persiapan penelitian yang perlu dilakukan data X yang didapatkan perlu dipersiapkan untuk penelitian, berikut tahapan yang perlu dilakukan.

Pelebelan Data

Pelebelan merupakan peroses memberikan label terhadap data berupa sentimen, positif, netral dan negatif (Liu, 2012). Pelabelan ini bertujuan untuk menghasilkan data yang berguna sebagai data pelatihan model LSTM dan BERT nantinya. Pada penelitian ini pelabelan dilakukan menggunakan *tool* dari Python yang bernama *Lexicon-VADER*.

VADER (*Valence Aware Dictionary and sentiment Reasoner*) adalah metode analisis sentimen berbasis *lexicon* yang dikembangkan khusus untuk

Bahasa Inggris. VADER mengandalkan kamus kata-kata sentimen (*sentiment lexicon*) yang berisi lebih dari 7.500 entri kata dan telah dipetakan ke skor polaritas, serta aturan gramatikal dan sintaksis yang umum digunakan dalam Bahasa Inggris, seperti penggunaan huruf kapital, tanda seru, konjungsi seperti “but”, dan juga emotikon (Hutto & Gilbert, 2014). Berikut langkah-langkah pelabelan dengan VADER.

Terjemahkan teks

Pada pelabelan dengan *lexicon* hanya memahami kata ber Bahasa Inggris maka proses terjemah ini penting proses ini menggunakan fungsi *translator* dari *library googletrans*. Namun, hal ini memiliki kelemahan, seperti hilangnya konteks budaya atau arti kata yang tidak tepat. Data yang sudah berbentuk teks Bahasa Inggris akan dilakukan analisa menggunakan *SentimentIntensityAnalyzer* dari *library NLTK* model *SentimentIntensityAnalyzer* merupakan model yang menggunakan kamus VADER untuk melakukan analisis sentimen. Setelah selesai akan menghasilkan kolom baru berupa sentimen (Hutto & Gilbert, 2014).`

Penanganan Imbalance Data

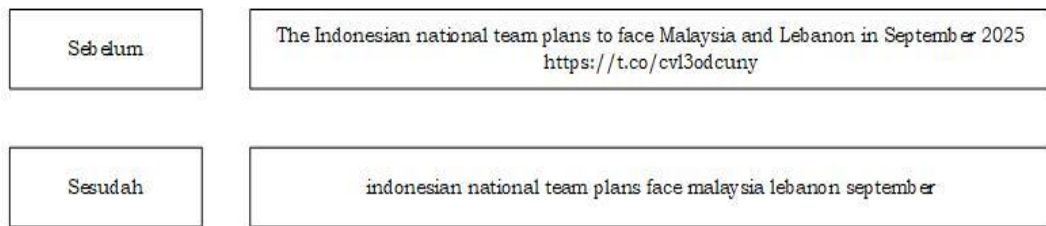
Imbalance data berpengaruh terhadap hasil klasifikasi karena model akan sulit untuk mempelajari metode yang dibangun, pada tahap ini peneliti melakukan *random undersampling* untuk menangani *imbalance* data dengan menyetarakan semua data dengan lebel minoritas dimana sebagian data mayoritas akan diapus tujuannya untuk mengurangi waktu *training* bisa memakan waktu yang lama, dan menghindari penumpukan data yang sama (Vluymans, 2019).

Preprocessing Data

Preprocessing atau prapemrosesan data merupakan tahap penting dalam pemodelan teks, khususnya saat menggunakan algoritma berbasis *deep learning* seperti LSTM dan BERT. *Preprocessing* bertujuan untuk membersihkan dan menyiapkan data teks agar sesuai dengan format *input* yang dibutuhkan oleh model, serta meningkatkan kualitas representasi data (Indurkha & Damerau, 2010).

Proses ini menggunakan *library re* dan untuk menghapus *noise*, dan karena data berikut beberapa hal yang dilakukan dalam *preprocessing* teks,

1. Mengkonversi teks menjadi huruf kecil.
2. Menghapus mension, hastag, dan URL.
3. Menghapus selain huruf, seperti angka, emoji dan tanda baca.
4. Menghapus *stopword*.

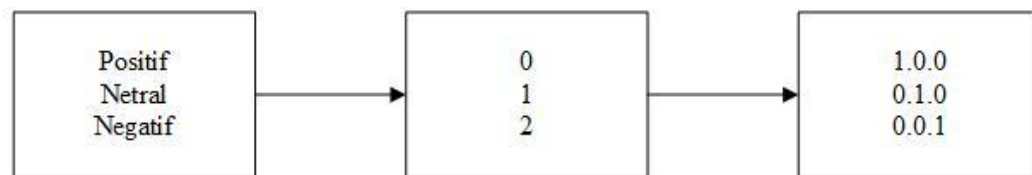


Gambar 5. Contoh Hasil Penerapan Membersihkan Teks

Tahap selanjutnya adalah melakukan *encode* label dan tokenisasi untuk mengubah teks menjadi angka agar bisa dilakukan penelitian.berikut beberapa proses *preprocessing*.

1. *Encode* Label

Encode label merupakan proses untuk mengubah label sentimen menjadi angka agar bisa dibaca sistem computer, proses ini terdapat dua tahap yang pertama mengubah menjadi angka biasa dan kemudian angka tersebut dilakukan *encode* kembali menjadi *one hot encoding* karena data berbentuk *multiclass*.



Gambar 6. Perubahan Data *Encoding*

2. Tokenisasi

Tokenisasi merupakan peroses memecah setiap kata dan memberikannya pada LSTM dan BERT unit-unit kata diubah menjadi format angka yang sudah tersedia pada fungsi *Tokenizer* untuk LSTM dn fungsi *BertTokenizer* untuk BERT (Jurafsky & Martin, 2021).

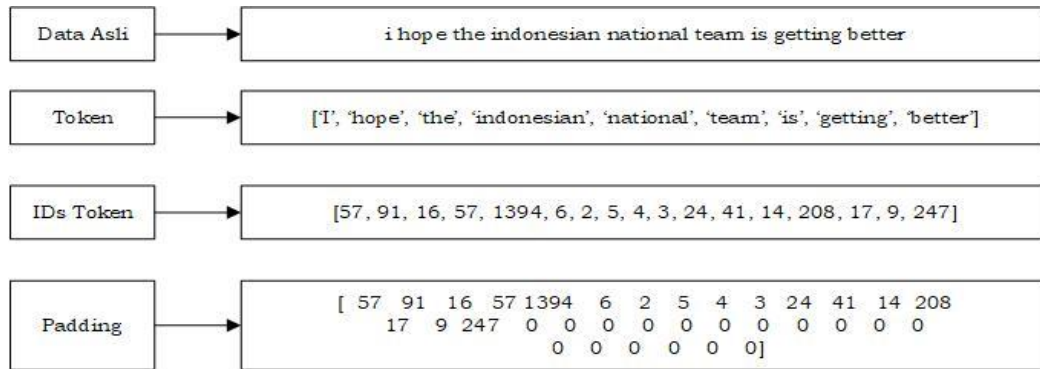
Tokenisasi LSTM

Tokenisasi menggunakan fungsi *tokenizer* dari *library keras* proses pemberian token pada menggunakan *tokenizer* berdasarkan dari jumlah kata yang paling sering muncul, semakin sering kata muncul maka token dari kata tersebut nilainya akan semakin kecil (Goldberg, 2017). Berikut langkah tokenisasi LSTM.

1. Data akan di proses *Tokenizer* untuk membuat kamus data dan mengubahnya menjadi angka berdasarkan seberapa sering kata muncul.
2. Setelah kamus dibuat, data akan diimplementasikan untuk diubah menjadi angka.

3. Kemudian data disimpan dalam *padding* untuk menyesuaikan dengan maksimal panjang dari data yang lain.
4. Token akan dibaca dan dipahami model LSTM.

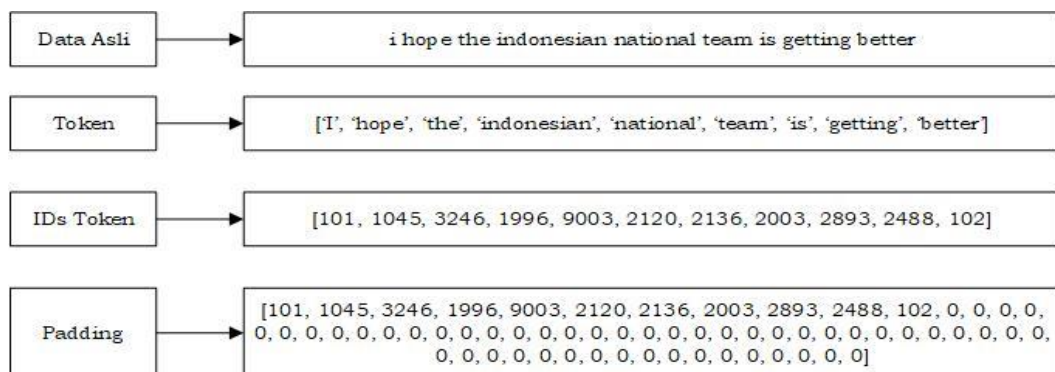
Berikut contoh penerapan tokenisasi Jika maksimal padding 34 pada Gambar 7.



Gambar 7. Contoh Penerapan Tokenisasi LSTM

Tokenisasi BERT

Tokenisasi menggunakan fungsi *BertTokenizer* dari *transformer* untuk tokenisasi model BERT dengan menggunakan tokenisasi dari *pretraining* '*bert-based-uncased*'. pada BERT terdapat *padding* tambahan yaitu [CLS] dan [SEP] dan bernilai [101] dan [102] dan token BERT tersebut dibungkus dengan fungsi *Dataset PyTorch* supaya bisa dibaca ketika menjalankan model BERT (Devlin et al., 2019). Contoh penerapan tokenisasi pada Gambar 8.



Gambar 8. Penerapan Tokenisasi BERT

3. Split Data

Split data sebenarnya bukan merupakan *preprocessing* namun pada tokenisasi LSTM memerlukan *tokenizer* yang belajar dari data *training* terlebih dahulu maka *split* data dicantumkan pada tahap ini. *Split* data merupakan proses membagi data menjadi dua dengan 80% data *training* dan 20% data *test* bertujuan untuk menjadikan 80% data sebagai data yang akan dilakukan

training untuk membangun model klasifikasi dan 20% data dijadikan sebagai data yang akan diuji terhadap model yang dibangun.

3.7 Metode Klasifikasi Data

Model yang digunakan untuk klasifikasi data pada penelitian ini adalah LSTM dan BERT dengan Python, penggunaan model ini akan diimplementasikan ketika data yang diambil dari proses *crawling* data X sudah diterjemahkan kedalam Bahasa Inggris untuk dilakukan pelabelan dengan *Lexicon VADER*, setelah data siap maka data akan dilakukan penyeimbangan data, *preprocessing*, pemodelan, *training*, dan pengujian.

Pemodelan Klasifikasi

Tahap ini dilakukan untuk membangun model klasifikasi dengan LSTM dan BERT, model ini berisi fungsi yang berasal dari LSTM dan BERT untuk melakukan klasifikasi dan menjadi acuan dalam *training*.

Dimulai dari LSTM:

1. Model LSTM dibangun dengan struktur *sequential*, dimana setiap *layer* pemodelan akan dijalankan dengan urutan.
2. Di dalam *layer* pertama adalah *embedding*, *embedding* disini bertugas untuk mengubah angka dari data menjadi vector.
3. *Layer* selanjutnya adalah LSTM, *layer* ini menjalankan algoritma dari LSTM dimana membaca, menangkap, dan mengingat data.
4. *Dense layer* ini memproses informasi dari LSTM dan memberikan *probabilitas* dari data.
5. Dan terakhir adalah *compile* yang berfungsi untuk menyiapkan model sebelum dilatih *training* seperti, mengukur kesalahan model, mengatur cara model belajar, memantau performa model selama pelatihan.

Selanjutnya pemodelan BERT. Pemodelan pada BERT menggunakan model *BertForSequenceClassification* yang merupakan model dari *library transformers* yang digunakan untuk klasifikasi data dan menggunakan *pre-trained* yang sudah ada yaitu *bert-base-uncased* yang mana didalamnya sudah terdapat lapisan-lapisan untuk melakukan klasifikasi.

Training

Training merupakan proses untuk melatih model untuk memahami pola atau hubungan dalam data supaya nantinya bisa membuat prediksi atau klasifikasi yang akurat terhadap data baru.

Proses *Training* Pada LSTM adalah sebagai berikut

1. Menjalankan model dengan data sudah disediakan.
2. Mengambil data latih dan Label dalam bentuk *one-hot encoder*.

3. 20% dari data digunakan untuk menguji performa model saat pelatihan berlangsung.
4. Dimana, Maksimum pelatihan selama 128 kali.
5. Setiap pelatihan model 64 data sekaligus akan dilatih pada setiap langkah.
6. Aktifkan *callback* yang berisi *EarlyStopping* dan *ReduceLROnPlateau* agar pelatihan lebih efisien.
7. Menggunakan *EarlyStopping* untuk menghentikan pelatihan jika validasi *loss* tidak membaik selama beberapa *epoch*.
8. Menggunakan *ReduceLROnPlateau* untuk menurunkan *learning rate* jika performa tidak membaik.

Kemudian *Training* pada BERT, berikut langkah-langkahnya:

1. Menyiapkan *TrainingArguments* yang digunakan untuk mengatur semua pengaturan *training* saat melatih model menggunakan *Trainer*. Terdapat beberapa aturan dalam *TrainingArguments* diantaranya adalah:
 - Mengatur folder tempat untuk menyimpan model hasil pelatihan dan *log* pelatihan, yang berguna untuk grafik atau pemantauan.
 - Mengatur ukuran *batch* saat model sedang dieuji.
 - Melakukan evaluasi pengukuran performa pada data test akan dilakukan setiap sekali pengujian dan model akan disimpan setiap selesai 1 kali pengujian.
 - Setelah semua pelatihan selesai, model terbaik akan otomatis digunakan.
2. Menyiapkan *Trainer* Komponen utama untuk melatih dan mengevaluasi model tanpa harus menulis *loop* pelatihan manual. *Trainer* juga mengatur semua elemen untuk proses pelatihan. Seperti
 - Model yang akan dilatih.
 - *TrainingArguments*.
 - Data *train* dan *test*.
 - Data *Tokenizer* BERT
 Semua elemen tersebut disimpan dalam *Trainer*
3. Menjalankan *trainer.train()*, merupakan perintah untuk memulai proses pelatihan model menggunakan *Trainer*

3.8 Prediksi dan Evaluasi Model

Pada tahap ini model LSTM dan BERT yang sudah dibangun dan dilatih digunakan untuk memprediksi *probabilitas* dari setiap baris pada data *test*. Setelah menjalankan melakukan prediksi selanjutnya hasil dari prediksi

dilakukan evaluasi. Evaluasi bertujuan untuk mengetahui kinerja model, mengukur kelebihan dan kekurangan model, dan juga menilai kualitas dari hasil prediksi dengan model LSTM dan BERT. Proses evaluasi model dilakukan dengan menggunakan fungsi *classification_report*, *confusion_matrix* untuk mendapatkan hasil *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Setelah itu divisualisasi dengan *matplotlib.pyplot* dan *ConfusionMatrixDisplay*.

3.9 Tahapan Visualisasi Data

Langkah selanjutnya adalah visualisasi data. Visualisasi data pada penelitian ini menggunakan *wordcloud*. Penggunaan *wordcloud* membuat visualisasi yang ditampilkan dibuat menarik dan informatif. Banyaknya kata dalam keseluruhan data mempengaruhi besar atau kecilnya kata tersebut pada gambar *wordcloud*. Semakin banyaknya kata tersebut, maka ukuran kata tersebut pada gambar *wordcloud* akan semakin besar, begitu pula sebaliknya. Semakin sedikitnya kata yang ada maka ukuran kata yang ditampilkan kecil.

IV. PEMBAHASAN

4.1 Pengumpulan Data

Untuk melakukan penelitian ini hal pertama yang harus dilakukan adalah mengumpulkan data yang diambil dari *tweet* pengguna aplikasi X dari Indonesia dengan kata kunci pencarian “PSSI” dan “Timnas Indonesia” yang dilakukan dengan teknik *crawling* data X melalui *tweet-harvest* v.2.6.1 yang dikembangkan Helmi Satria. Teknik ini menggunakan metode *scrolling* otomatis dan mengambil data dari hasil *scrolling*. Untuk melakukan *crawling* dengan *tweet-harvest* penulis perlu memasukkan *X_auth_token* yang didapatkan dari akun X yang akan dijadikan sebagai sumber data, setelah itu dilanjutkan dengan memasang program *tweet-harvest* beserta *node.js*, kode program untuk menyiapkan *tweet-harvest* bisa dilihat pada Gambar 9.

```
#Twitter Auth Token
twitter_auth_token =
'be5255cab0f1677dbd88e6f9a3b301d8075cfc91'

# Menginstal pustaka Python 'pandas' yang diperlukan untuk
analisis data
!pip install pandas

# Memperbarui daftar paket di sistem agar instalasi
selanjutnya menggunakan versi terbaru
!sudo apt-get update

# Menginstal dependensi yang diperlukan untuk mengatur
repositori Node.js
!sudo apt-get install -y ca-certificates curl gnupg

# Membuat direktori untuk menyimpan keyring (penyimpanan kunci
enkripsi) Node.js
!sudo mkdir -p /etc/apt/keyrings

# Mengunduh dan menyimpan GPG key Node.js ke dalam keyring
untuk memastikan keamanan instalasi
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg

# Menambahkan repositori resmi Node.js versi 20 ke dalam
sources list sistem
!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
```

```

https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" |
sudo tee /etc/apt/sources.list.d/nodesource.list

# Memperbarui daftar paket lagi setelah menambahkan repositori
Node.js
!sudo apt-get update

# Menginstal Node.js dari repositori yang telah ditambahkan
!sudo apt-get install nodejs -y

# Menampilkan versi Node.js yang telah terinstal untuk
verifikasi
!node -v

```

Gambar 9. Kode Persiapan Proses *Crawling*

Setelah persiapan selesai maka tahap selanjutnya adalah dengan membuat nama file csv yang akan digunakan untuk menyimpan data, kemudian pada tahap selanjutnya adalah memasukkan kata kunci yang diinginkan beserta batas waktu kapan *tweet* tersebut dibuat dan juga data yang diambil juga dibatasi untuk pengguna Bahasa Indonesia dapat dilihat Pada Gambar 10.

```

# Menentukan nama file output untuk menyimpan data hasil crawl
filename = 'data_mentah.csv'

# Menentukan kata kunci pencarian untuk Tweet yang berbahasa
Indonesia,
# mencakup topik "PSSI dan Timnas Indonesia" dari tanggal 27
Juni 2024 hingga 1 Mei 2025
search_keyword = 'PSSI and Timnas Indonesia since:2024-06-27
until:2025-05-01 lang:id'

# Menetapkan batas jumlah tweet yang ingin diambil
limit = 4000

# Menjalankan tool tweet-harvest menggunakan NPX untuk
mengambil tweet berdasarkan parameter yang ditentukan:
# -o: nama file output
# -s: kata kunci pencarian
# --tab "LATEST": mengambil tweet terbaru (bukan populer atau
top)
# -l: jumlah maksimum tweet
# --token: token autentikasi Twitter (harus sudah disiapkan
sebelumnya)
!npx -y tweet-harvest@2.6.1 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}

```

Gambar 10. Kode Program Melakukan *Harvest-Tweet*

Pada penelitian ini data yang diambil dibatasi 4000 *tweet* karna proses *Crawling* yang memakan waktu lama dengan tujuan untuk mengulang kembali proses *Crawling* dengan pedoman tanggal terakhir data *tweet* yang masuk, namun hasil dari proses *crawling* data X diperoleh data X hanya sebanyak 3981 dan tidak mencapai 4000 data dengan data terakhir yang didapatkan adalah pada tanggal 30 April 2025. Didalam data tersebut mencakup banyak variabel seperti *tweet_id*, *text*, *username*, *created_at*, *retweet_count*, *like_count*, dan *lang* tanggal, *username*, lokasi, isi *tweet* yang berhubungan dengan PSSI dan Timnas dan juga beberapa variabel lainnya dan disimpan dalam format csv. Berikut data yang terimpan dalam format csv dari hasil *crawling* dapat dilihat pada Gambar 11.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
conversation_id	created_at	favo	full_text	id_str	image_url	in_reply_to	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id	username
1	1.92E+18 Wed Apr 30 23:45:00 +0000 2025	0	Timnas Indonesia Berencana Hadapi Malaysia dan	1.92E+18			in		1	0	0	https://x.com/47274731		
2	1.92E+18 Wed Apr 30 23:19:34 +0000 2025	0	Ole Romeny Blak-blakan Mau Bela Timnas Indone	1.92E+18			in		0	0	0	https://x.com/47274731		
3	1.92E+18 Wed Apr 30 23:15:00 +0000 2025	0	Asisten Pelatih Timnas Indonesia Denny Landzaat	1.92E+18			in		0	0	0	https://x.com/47274731		
4	1.92E+18 Wed Apr 30 22:10:10 +0000 2025	0	Timnas putri Indonesia umumkan 26 nama untuk F	1.92E+18			in		0	0	0	https://x.com/1.4E+09		
5	1.92E+18 Wed Apr 30 22:00:01 +0000 2025	0	Pratama Arhan Mulai Kehilangan Tempat di Timna	1.92E+18			in		0	0	0	https://x.com/47274731		
6	1.92E+18 Wed Apr 30 21:00:01 +0000 2025	1	5 Pemain Naturalisasi Tambahan yang Dijagokan D	1.92E+18			in		0	0	0	https://x.com/47274731		
7	1.92E+18 Wed Apr 30 20:24:55 +0000 2025	0	@idextratime Uda kayak pertahanan timnas Indon	1.92E+18			idextratin		0	0	0	https://x.com/1.07E+08		
8	1.92E+18 Wed Apr 30 20:09:51 +0000 2025	0	Detail best offer #betis1 Kamis 1 Mei 2025 Timnas	1.92E+18	https://pbs.twimg.c		in		1	2	0	https://x.com/1.25E+18		
9	1.92E+18 Wed Apr 30 19:53:12 +0000 2025	0	Alex Pastoor Ngamuk Agenda Lawan Timnas Indon	1.92E+18			in		0	0	0	https://x.com/1.82E+18		
10	1.92E+18 Wed Apr 30 19:27:37 +0000 2025	0	Detail best offer #betis11 Kamis 1 Mei 2025 Timnas	1.92E+18	https://pbs.twimg.c		in		1	1	0	https://x.com/1.25E+18		
11	1.92E+18 Wed Apr 30 19:23:34 +0000 2025	17	@Jerseyforum Uda dipake asisten timnas Indon	1.92E+18	https://pt.jerseyfori		in		1	0	2	https://x.com/1.57E+18		
12	1.92E+18 Wed Apr 30 19:00:00 +0000 2025	1	Timnas Indonesia Jadi Tuan Rumah Babak Keempa	1.92E+18			in		0	0	0	https://x.com/47274731		
13	1.92E+18 Wed Apr 30 18:00:01 +0000 2025	0	Patrick Kluivert Panggil Ciro Alves ke Timnas Indor	1.92E+18			in		0	0	0	https://x.com/47274731		
14	1.92E+18 Wed Apr 30 17:40:04 +0000 2025	0	[Hemat hingga 58%] Rasakan kenyamanan dan gay	1.92E+18			in		0	0	0	https://x.com/3.11E+09		
15	1.92E+18 Wed Apr 30 17:06:44 +0000 2025	1	Tertarik mengetahui gaji pemain Timnas U-20 Ind	1.92E+18			in		0	0	0	https://x.com/1.90E+18		
16	1.92E+18 Wed Apr 30 16:30:57 +0000 2025	0	Anak pelatih Timnas Indonesia Shane Kluivert den	1.92E+18	https://pbs.twimg.c		in		0	0	0	https://x.com/1.88E+18		
17	1.92E+18 Wed Apr 30 16:17:48 +0000 2025	0	Apa yg kalian cari? Ngebon pemain mengecek pelu	1.92E+18			in		0	0	0	https://x.com/1.8E+08		
18	1.92E+18 Wed Apr 30 16:16:01 +0000 2025	0	Pratama arhan punggawa skuad garuda timnas ind	1.92E+18	https://pbs.twimg.c		in		0	0	0	https://x.com/1.42E+18		
19	1.92E+18 Wed Apr 30 16:13:25 +0000 2025	0	Cek Jersey Bola Timnas Indonesia 2025 Dryfit Dewi	1.92E+18	https://pbs.twimg.c		in		0	0	0	https://x.com/1.82E+18		
20	1.92E+18 Wed Apr 30 16:06:14 +0000 2025	1	Cek Jersey Timnas Indonesia 2025 Grade Ori Dryfit	1.92E+18	https://pbs.twimg.c		in		0	0	0	https://x.com/1.82E+18		
21	1.92E+18 Wed Apr 30 15:16:26 +0000 2025	0	Erick Thohir Menolak Gagasan Jadi Ketua Umum PS	1.92E+18			in		0	0	0	https://x.com/1.35E+09		
22	1.92E+18 Wed Apr 30 15:01:09 +0000 2025	71	1.07E+08 Bersiap menyambut masa depan yang	1.92E+18	https://pbs.twimg.c		in		0	1	6	https://x.com/2.36E+09		
23	1.92E+18 Wed Apr 30 15:00:47 +0000 2025	0	@butoable @idextratime @Plettigoal Ok berarti X	1.92E+18			butoable		0	1	0	https://x.com/1.03E+18		
24	1.92E+18 Wed Apr 30 15:00:02 +0000 2025	0	Klasemen Sementara Calon Tuan Rumah Babak Ke	1.92E+18			in		0	0	0	https://x.com/47274731		

Gambar 11. Hasil Pengumpulan Data dengan *Tweet-Harvest*

4.2 Pelabelan Data

Sebelum pelebelan karna *tool* analisis sentimen saat ini rata-rata menggunakan pelatihan kosakata dalam Bahasa Inggris maka perlu dilakukan terjemah kata.

Menerjemahkan teks

Menerjemahkan teks bertujuan agar diharapkan hasil dari pelabelan yang dilakukan menggunakan *tool* bisa menghasilkan akurasi yang lebih baik karena didalam *tool* kosakata dalam Bahasa Inggris jauh lebih banyak dibandingkan bahasa lain. Untuk menerjemahkan kata langkah langkah yang

harus dilakukan adalah menginstall *library googletrans*. Setelah *library* terpasang maka selanjutnya adalah menuliskan kode program untuk menerjemahkan teks menjadi Bahasa Inggris seperti Gambar 16. Dan hasilnya pada Tabel 2.

```
# Mengimpor modul Translator dari pustaka googletrans
from googletrans import Translator

# Membuat objek Translator untuk melakukan penerjemahan
translator = Translator()

# Fungsi untuk menerjemahkan teks secara aman dari Bahasa
Indonesia ke Bahasa Inggris
def translate_text(text):
    try:
        # Menerjemahkan teks dengan menentukan bahasa sumber
        ('id') dan tujuan ('en')
        return translator.translate(str(text), src='id',
dest='en').text
    except:
        # Jika terjadi error (misalnya koneksi atau teks
kosong), kembalikan teks aslinya
        return text

# Menerapkan fungsi translate_text ke kolom 'full_text' dalam
DataFrame
# Hasil terjemahan disimpan ke kolom baru bernama
'hasil_terjemah'
df['hasil_terjemah'] = df['full_text'].apply(translate_text)
```

Gambar 12. Teks Bahasa Indonesia Ke Bahasa Inggris

Tabel 2. Hasil Terjemah

No	Tweet	Terjemahan
1	<i>Timnas Indonesia Berencana Hadapi Malaysia dan Lebanon pada September 2025</i> https://t.co/cvL3OdcUNy	The Indonesian national team plans to face Malaysia and Lebanon in September 2025 https://t.co/cvL3OdcunY
2	<i>Ole Romeny Blak-blakan Mau Bela Timnas Indonesia Berkat Fans Garuda</i> https://t.co/81kaqJYTbi	Ole Romeny, I want to defend the Indonesian national team thanks to Garuda fans https://t.co/81Kagjytbi
3	<i>Asisten Pelatih Timnas Indonesia Denny Landzaat Dibidik Jadi Pelatih Willem II</i> https://t.co/2xhXWw6z1H	Indonesian national team assistant coach Denny Landzaat is targeted to be Willem II coach https://t.co/2xhxww6z1h

Sumber : Hasil Olah Data, 2025

Setelah diterjemahkan teks akan disimpan menjadi file csv baru dengan nama `data_analisa_baru.csv` yang bertujuan untuk mempermudah proses selanjutnya. Kemudian penulis melakukan pelabelan dengan menggunakan *tool* dari Python yaitu VADER, dan perlu untuk memasang library `nlTK` seperti. *library nlTK* digunakan untuk melakukan proses pelabelan dengan *tool* VADER *lexicon* yang tersedia pada NLTK, dan visualisasi dari proses pelabelan bisa dilihat pada gambar 13.

```
# Mengimpor pustaka yang diperlukan
import nltk # Pustaka untuk pemrosesan bahasa alami
from nltk.sentiment import SentimentIntensityAnalyzer # Kelas
untuk analisis sentimen
import matplotlib.pyplot as plt # Pustaka untuk visualisasi
grafik

# Mengunduh lexicon VADER yang diperlukan
nltk.download('vader_lexicon')

# Inisialisasi objek SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()
# Fungsi untuk menentukan sentimen berdasarkan skor komposit
def get_sentiment(text):
    score = sia.polarity_scores(text) # Menghitung skor
    sentimen
    compound = score['compound'] # Mengambil skor komposit
    if compound >= 0.05:
        return 'positive' # Sentimen positif
    elif compound <= -0.05:
        return 'negative' # Sentimen negatif
    else:
        return 'neutral' # Sentimen netral

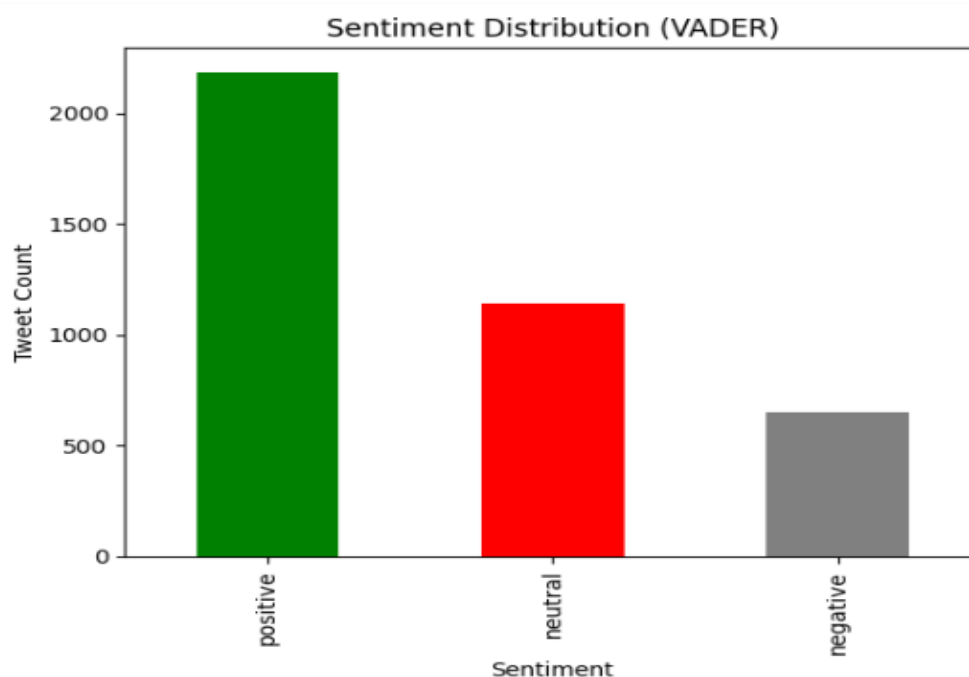
# Menerapkan fungsi get_sentiment pada kolom 'hasil_terjemah'
DataFrame df
df['sentiment'] = df['hasil_terjemah'].apply(get_sentiment)
# Visualisasi distribusi sentimen
df['sentiment'].value_counts().plot(kind='bar',
color=['green', 'red', 'gray'])
plt.title('Sentiment Distribution (VADER)') # Judul grafik
plt.xlabel('Sentiment') # Label sumbu x
plt.ylabel('Tweet Count') # Label sumbu y
plt.show() # Menampilkan grafik
```

Gambar 13. Kode Pelabelan dengan VADER

Tabel 3. Hasil Dari Pelabelan VADER

No	Tweet	Sentimen
1	<i>The Indonesian national team plans to face Malaysia and Lebanon in September 2025</i> https://t.co/cv13odcuny	<i>Neutral</i>
2	<i>Ole Romeny, I want to defend the Indonesian national team thanks to Garuda fans</i> https://t.co/81Kagjytbi	<i>Positive</i>
3	<i>Not as strong as how strong the Indonesian national team refused players who had played in 2.bundesliga .. fans2 like this usually the big head.</i>	<i>Negative</i>

Sumber : Hasil Olah Data, 2025

**Gambar 14.** Visualisasi Hasil Pelabelan dengan VADER

Dari Pelabelan dengan VADER didapatkan hasil data sentimen positif sebanyak 2185, negatif 652, dan netral sebanyak 1143 data. Karena data pelabelan yang didapat tidak seimbang antara lebel satu dan lainnya maka untuk mempermudah proses pengujian model maka perlu dilakukan proses penanganan *imbalance* data.

4.3 Penanganan *Imbalance* Data

Imbalance data merupakan kondisi didalam dataset klasifikasi, dimana distribusi antar kelas tidak merata atau tidak seimbang. Sebagai contoh jumlah dari satu kelas jauh lebih banyak dibandingkan kelas lainnya. Hal ini dapat

menyebabkan bias model terhadap kelas mayoritas, sehingga performa model pada kelas minoritas menurun drastis (He & Garcia, 2009).

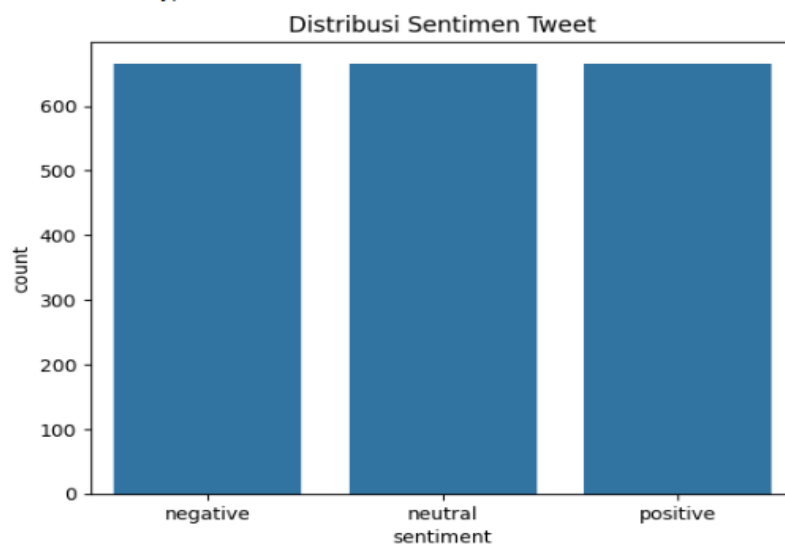
Dan metode dilakukan pada penelitian adalah menggunakan pandas dengan *random undersampling* dengan mengurangi jumlah data dari kelas mayoritas. Hasil dari *undersampling* bisa dilihat pada Gambar 16.

```
# Tujuannya untuk menyeimbangkan jumlah data antar kelas
min_count = df['sentiment'].value_counts().min()

# Menyeimbangkan dataset dengan cara mengambil sampel acak
# dari setiap kelas sebanyak 'min_count'
# Sehingga semua kelas memiliki jumlah data yang sama

df_balanced = df.groupby('sentiment').apply(lambda x:
x.sample(min_count)).reset_index(drop=True)
```

Gambar 15. Kode Program *Undersampling*



Gambar 16. Hasil dari *Random Undersampling*

Setelah data didapatkan data yang seimbang maka tahap selanjutnya adalah melakukan penyiapan data untuk pelatihan model LSTM dan BERT.

4.4 Preprocessing Data

Data yang sudah dikumpulkan perlu dilakukan tahap *preprocessing* untuk menghilangkan hal yang tidak diperlukan pada teks dan tidak ada hubungannya dalam penelitian sehingga teks bisa siap untuk digunakan dalam penelitian. Tahap ini meliputi menghilangkan *noise* pada teks dan menerjemahkan teks setelah melakukan *preprocessing* data siap untuk dilakukan penelitian tahap selanjutnya.

Membersihkan Teks

Membersihkan teks bertujuan untuk menghilangkan elemen yang tidak diperlukan dalam teks seperti *url*, *mansion/tag*, dan simbol untuk mempermudah proses pelabelan. Data yang sudah dilabelkan selanjutnya dilakukan proses membersihkan teks dengan menggunakan *library re* dan fungsi *stopword* dari *nltk.corpus*. *Re* merupakan modul yang sangat berguna dalam pengolahan teks pada Python, seperti menghapus *URL* dan tanda baca dari teks terutama pada data yang memerlukan deteksi pola atau format seperti *email*. (Mckinney, 2017). Selain menggunakan modul *re*, juga menggunakan *NLTK* karena perlu untuk menghilangkan *stopwords*. *Stopwords* merupakan kata-kata yang tidak terlanlu memiliki banyak arti. Proses ini memerlukan modul dari *NLTK* untuk menghilangkan *stopwords*. Kode program untuk membersihkan teks, dan menghilangkan *stopwords* pada Gambar 17. Dan hasilnya pada Tabel 4.

```
import re
import nltk
import pandas as pd
from nltk.corpus import stopwords

# Unduh stopwords bahasa Inggris
nltk.download('stopwords')

# Set stopwords bahasa Inggris
stop_words = set(stopwords.words('english'))

# Fungsi: Pembersihan teks (lowercase, hapus URL, mention,
# hashtag, simbol)
def clean_text(text):
    text = text.lower()
    text = re.sub(r"http\S+", "", text)           # Hapus URL
    text = re.sub(r"@w+", "", text)               # Hapus mention
    (@user)
    text = re.sub(r"#w+", "", text)               # Hapus hashtag
    (#tag)
    text = re.sub(r"[^a-zA-Z\s]", "", text)        # Hapus
    karakter non-alfabet
    text = re.sub(r"\s+", " ", text).strip()      # Hapus spasi
    berlebih
    return text

# Fungsi: Stopword removal
def remove_stopwords_en(text):
    words = text.split()
```

```

    filtered = [word for word in words if word not in
stop_words]
    return " ".join(filtered)

# 1. Bersihkan teks
df['clean_text'] = df['hasil_terjemah'].apply(clean_text)

# 2. Hapus stopwords
df['clean_text'] = df['clean_text'].apply(remove_stopwords_en)

```

Gambar 17. Kode Program Untuk Membersihkan Teks

Tabel 4. Hasil Dari Membersihkan Data

No	<i>Tweet</i>	Hasil Pembersihan
1	The Indonesian national team plans to face Malaysia and Lebanon in September 2025 https://t.co/cvl3odcuny	indonesian national team plans face malaysia lebanon september
2	Ole Romeny, I want to defend the Indonesian national team thanks to Garuda fans https://t.co/81Kagjytbi	ole romeny want defend indonesian national team thanks garuda fans
3	Indonesian national team assistant coach Denny Landzaat is targeted to be Willem II coach https://t.co/2xhxww6z1h	indonesian national team assistant coach denny landzaat targeted willem ii coach

Sumber : Hasil Olah Data, 2025

4.5 Encode Label

Encode Label merupakan proses mengubah label *string* positif, netral, negatif ke angka [0, 1, 2] menggunakan perintah *LabelEncoder* dari *library scikit-learn*. Kode dari proses *encode* bisa dilihat pada Gambar 18. Dan hasilnya pada Tabel 5.

```

# Mengimpor LabelEncoder dari scikit-learn untuk mengubah
label kategorikal menjadi angka
from sklearn.preprocessing import LabelEncoder

# Membuat objek LabelEncoder
label_encoder = LabelEncoder()

# Mengubah nilai pada kolom 'sentiment' menjadi angka, dan
menyimpannya ke kolom baru 'label_encoded'
# Contoh: 'positif' → 2, 'netral' → 1, 'negatif' → 0 (urutan
tergantung isi data)
df['label_encoded'] =
label_encoder.fit_transform(df['sentiment'])

# Menyimpan jumlah kelas unik dalam variabel num_classes
(berguna untuk keperluan pelatihan model)
num_classes = len(label_encoder.classes_)

```

Gambar 18. Kode Program *Encode*

Tabel 5. Hasil *Encode*

Sentimen	Encode
Negativ	0
Neutral	1
Positive	2

Sumber : Hasil Olah Data, 2025

4.6 Split Data

Pada Tahap ini merupakan proses untuk membagi data menjadi 80% untuk *training* dan 20% untuk *testing*. Dengan menggunakan perintah *train_test_split* dari *sklearn.model_selection*. Kode *split* data pada Gambar 19.

```
# Mengimpor fungsi train_test_split dari scikit-learn untuk
membagi dataset menjadi data latih dan data uji
from sklearn.model_selection import train_test_split

# Membagi data menjadi data latih (80%) dan data uji (20%)
dengan mempertahankan proporsi label (stratifikasi)
train_texts, test_texts, train_labels, test_labels =
train_test_split(

    df['clean_translated_text'],      # Fitur input berupa teks
yang telah dibersihkan dan diterjemahkan
    df['label_encoded'],              # Label target dalam
bentuk numerik

    test_size=0.2,                   # 20% data digunakan
sebagai data uji
    random_state=42,                 # Seed untuk memastikan
hasil pembagian data tetap sama jika dijalankan ulang
    stratify=df['label_encoded']
    # Memastikan distribusi label di data latih dan data uji
tetap proporsional

)
```

Gambar 19. Kode Program Untuk *Split* Data

Pada tahap split data didapatkan hasil berupa 1564 data *training* dan 392 data *testing*.

4.7 Tokenisasi

Tokenisasi merupakan peroses memecah kalimat menjadi kata perkata dan juga mengubah menjadi bentuk angka. Pada tahap ini terdapat perbedaan antara LSTM dan BERT dimana LSTM menggunakan perintah *Tokenizer* dari

library tensorflow.keras.preprocessing.text, sementara BERT menggunakan *BertTokenizer* dari *library hugging face transformers*.

Tokenisasi LSTM

Proses tokenisasi LSTM perlu menggunakan fungsi *Tokenizer* dari modul *tensorflow.keras.preprocessing.text* yang berfungsi Mengubah teks menjadi urutan angka. Dan menyusun kamus kata berdasarkan dari frekuensi kata yang muncul dalam data. LSTM juga menggunakan fungsi *pad_sequences* dari *tensorflow.keras.preprocessing.sequence*. *Pad_sequences* yang digunakan untuk menyeragamkan panjang setiap *input* dengan menambahkan *padding* bernilai nol ke awal atau akhir juga bisa memotong *sequence* yang terlalu panjang. Karena LSTM membutuhkan semua *input* harus memiliki panjang yang sama. Berikut kode penerapan tokenisasi pada LSTM pada Gambar 20. Dan hasilnya pada Gambar 21.

```
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import
pad_sequences
import numpy as np

tokenizer = Tokenizer(num_words=10000, oov_token="<OOV>") # 1.
Tokenizer dengan num_words besar dan oov_token
tokenizer.fit_on_texts(train_texts) # X_train adalah list of
text (teks mentah)

X_train_seq = tokenizer.texts_to_sequences(train_texts) #
Tokenisasi
oov_count = sum([seq.count(1) for seq in X_train_seq]) # 2.
Hitung jumlah token <OOV> (biasanya token 1)

lengths = [len(seq) for seq in X_train_seq] # 3. Analisis
panjang input untuk menentukan max_len
print("Panjang rata-rata kalimat:", np.mean(lengths))
print("Panjang 90% kalimat:", np.percentile(lengths, 90))

max_len = int(np.percentile(lengths, 90)) # Gunakan panjang
ke-90 persentil sebagai max_len
print(f"max_len yang disarankan: {max_len}")

X_train_pad = pad_sequences(X_train_seq, maxlen=max_len,
padding='post', truncating='post') # Padding sequence
```

Gambar 20. Gambar Kode Program Tahap Tokenisasi

Teks asli: check the indonesian national team soccer jersey price of rp
 Tokenisasi : [7, 10, 4, 3, 20, 11, 15, 26, 32, 6, 2, 12, 8, 7, 121, 131, 319, 23, 65, 110, 110, 282, 1917]
 Padding : [7 10 4 3 20 11 15 26 32 6 2 12 8 7
 121 131 319 23 65 110 110 282 1917 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0
 0 0]

Gambar 21. Contoh Hasil Dari Tokenisasi pada LSTM

Sebelum melakukan *training* pada LSTM perlu tahap *One-Hot Encoding*. *One-hot encoding* yang digunakan untuk mengubah label kategori (seperti 0, 1, 2) menjadi vektor biner agar bisa diproses dengan baik oleh model klasifikasi *deep learning*, terutama saat menggunakan fungsi *softmax*. *Softmax* adalah fungsi yang digunakan pada *output layer model deep learning* untuk klasifikasi multi-kelas (Yoshua Bengio et al., 2015). Proses ini menggunakan fungsi *to_categorical* dari Keras kode pada tahap ini seperti Gambar 22, dan hasil dari penerapan *One-Hot-Encodig* pada Tabel 6.

```
# Mengimpor fungsi to_categorical dari TensorFlow Keras
# Fungsi ini digunakan untuk mengubah label numerik menjadi
format one-hot encoding
from tensorflow.keras.utils import to_categorical

# Mengubah label pelatihan (train_labels) menjadi bentuk one-
hot encoding
# Misalnya label 2 akan menjadi [0, 0, 1] jika ada 3 kelas

y_train_cat = to_categorical(train_labels,
num_classes=num_classes)

# Mengubah label pengujian (test_labels) menjadi bentuk one-
hot encoding

y_test_cat = to_categorical(test_labels,
num_classes=num_classes)
```

Gambar 22. Kode Program Penerapan *One-Hot-Encodig*:

Tabel 6. Hasil Penerapan *One-Hot-Encoding*

<i>One-Hot-Encoding</i>	<i>Encode</i>
[1. 0. 0.]	0
[0. 1. 0.]	1
[0. 0. 1.]	2

Sumber : Hasil Olah Data, 2025


```

    # Konstruktor (__init__) untuk menyimpan hasil tokenisasi
    (encodings) dan label (labels)
    def __init__(self, encodings, labels):
        self.encodings = encodings # dictionary: input_ids,
        attention_mask, dll
        self.labels = labels # list label untuk setiap
        sampel
        # Method __getitem__ akan mengambil satu item berdasarkan
        indeks (idx)
        def __getitem__(self, idx):
            # Mengambil input sesuai index dan mengubahnya ke
            tensor PyTorch
            item = {k: torch.tensor(v[idx]) for k, v in
            self.encodings.items()}
            # Menambahkan label ke dalam item (juga sebagai
            tensor)
            item['labels'] = torch.tensor(self.labels[idx])
            return item # item akan berisi input_ids,
            attention_mask, dan labels dalam bentuk tensor
            # Method __len__ untuk mengembalikan panjang dataset
            (jumlah sampel)
            def __len__(self):
                return len(self.labels)
            # Membuat dataset PyTorch dari data latih (train)
            train_dataset = TweetDataset(train_encodings,
            list(train_labels))
            # Membuat dataset PyTorch dari data uji (test)
            test_dataset = TweetDataset(test_encodings, list(test_labels))

```

Gambar 25. Kode Penerapan *Dataset Class Pytech*

4.8 Penerapan Model Klasifikasi

Setelah melakukan tokenisasi selanjutnya adalah membangun model klasifikasi yang akan diuji dari LSTM dan BERT

Pemodelan LSTM

LSTM menggunakan model *Sequential* yang merupakan model dari Keras yang memungkinkan untuk membangun *neural network* secara berurutan. Selain itu pada tahap ini perlu untuk menggunakan beberapa lapisan *neural network* diantaranya *Embedding* yang berfungsi untuk representasi kata, LSTM berfungsi untuk menangkap urutan kata dari urutan waktu *Dropout* berfungsi untuk mencegah *overfitting* dan *Dense* merupakan lapisan untuk klasifikasi. Kode pemodelan LSTM bisa dilihat pada Gambar 26.

```

# Mengimpor modul untuk membangun model neural network
menggunakan Keras
from tensorflow.keras.models import Sequential

```

```

from tensorflow.keras.layers import Embedding, LSTM, Dropout,
Dense

# Membangun model sekuensial (berlapis satu per satu)
model = Sequential([
    # Layer embedding untuk mengubah token integer menjadi
    representasi vektor berdimensi 128
    # input_dim = jumlah kata unik (vocab size), output_dim =
    dimensi vektor, input_length = panjang input tetap
    Embedding(input_dim=10000, output_dim=128,
    input_length=max_len),

    # Layer LSTM untuk menangkap informasi urutan dalam teks
    (recurrent layer)
    # return_sequences=False berarti hanya mengambil output
    terakhir, cocok untuk klasifikasi
    LSTM(128, return_sequences=False),

    # Dropout layer untuk mengurangi overfitting dengan
    membuang 50% neuron secara acak saat training
    Dropout(0.5),

    # Dense layer (fully connected) dengan 64 neuron dan
    aktivasi ReLU
    Dense(64, activation='relu'),

    # Output layer dengan neuron sebanyak jumlah kelas dan
    aktivasi softmax untuk klasifikasi multi-kelas
    Dense(num_classes, activation='softmax')
])
# Menyusun model dengan:
# - loss function: categorical_crossentropy (karena one-hot
    encoding digunakan)
# - optimizer: Adam (algoritma optimisasi populer)
# - metrics: akurasi untuk evaluasi kinerja
model.compile(loss='categorical_crossentropy',
    optimizer='adam', metrics=['accuracy'])
# Menampilkan ringkasan arsitektur model
model.summary()

```

Gambar 26. Kode Pemodelan LSTM

Pemodelan BERT

BERT membangun pemodelan dengan *BertForSequenceClassification* yang merupakan model yang sudah disesuaikan untuk tugas klasifikasi teks dari *transformers*. Pemodelan disesuaikan dengan model BERT yang sudah dilatih sebelumnya didalam *bert-base-uncased* yang merupakan versi BERT

standar dari Google dan menyesuaikannya untuk klasifikasi teks. Model BERT untuk tugas klasifikasi teks mengambil model BERT *pretrained* dan menyesuaikan *output* klasifikasi dengan jumlah label. Kode perintah pemodelan pada Gambar 27.

```
# Mengimpor model BERT untuk tugas klasifikasi dari library
Transformers (Hugging Face)
from transformers import BertForSequenceClassification
# Memuat model BERT pre-trained ('bert-base-uncased') yang
disesuaikan untuk klasifikasi teks
# - num_labels menunjukkan jumlah kelas yang akan diprediksi
oleh model
# - Model ini secara otomatis menambahkan layer klasifikasi di
atas arsitektur BERT
model = BertForSequenceClassification.from_pretrained(
    'bert-base-uncased',                # Model BERT
    pre-trained dengan lowercase text
    num_labels=len(label_encoder.classes_) # Menentukan
jumlah kelas target (misalnya: positif, netral, negatif)
)
```

Gambar 27. Kode Program untuk Pemodelan BERT

4.9 Training Data

Setelah melakukan pemodelan maka tahap selanjutnya adalah melakukan *training*. *Training* adalah proses melatih model agar bisa belajar dari data, sehingga mampu memprediksi atau mengklasifikasikan data baru dengan benar. Berikut kode penerapan *training* pada model LSTM pada Gambar 28.

```
history = model.fit(
    X_train_pad,      # Data teks yang sudah diproses dan
dipadatkan (padding)
    y_train_cat,      # Label dalam bentuk one-hot encoding
    validation_split=0.2, # 20% data training dipakai sebagai
validasi
    epochs=20,        # Model dilatih sebanyak 20 kali
iterasi terhadap seluruh data
    batch_size=32      # Jumlah sampel per batch saat
training
)
```

Gambar 28. Kode Penerapan Training Model LSTM

Kemudian *training* pada model BERT. *Training* model BERT pada tahap ini memerlukan beberapa langkah diantaranya sebagai berikut, yang pertama *training args* berperan untuk mengkonfigurasi perintah *training*. Tahap ini menggunakan fungsi *TrainingArguments* dari library *transformers*. Kode program *training args* pada Gambar 29.

```

# Mengimpor TrainingArguments dari library Hugging Face
Transformers
# TrainingArguments digunakan untuk mengatur parameter
pelatihan model
from transformers import TrainingArguments

# Menentukan konfigurasi/parameter untuk proses pelatihan
model
training_args = TrainingArguments(
    output_dir='./results',          # Direktori untuk
    # Menyimpan model yang telah dilatih
    num_train_epochs=3,              # Jumlah epoch
    # (putaran pelatihan penuh terhadap seluruh data)
    per_device_train_batch_size=8,    # Ukuran batch
    # saat pelatihan untuk setiap perangkat (misalnya per GPU)
    per_device_eval_batch_size=8,     # Ukuran batch
    # saat evaluasi untuk setiap perangkat
    eval_strategy="epoch",            # Melakukan
    # evaluasi setiap selesai satu epoch
    save_strategy="epoch",            # Menyimpan model
    # setiap selesai satu epoch
    logging_dir='./logs',             # Direktori untuk
    # menyimpan log pelatihan
    load_best_model_at_end=True        # Secara otomatis
    # memuat model terbaik (berdasarkan evaluasi) setelah training
    # selesai
)

```

Gambar 29. Kode Program *Training Args*

Selanjutnya *trainer*, *trainer* berperan untuk menyederhanakan *training* dan evaluasi, tahap ini menggunakan fungsi *Trainer* dari *transformers*, lalu *DataCollatorWithPadding* merupakan komponen yang bisa otomatis menambahkan *padding* ke *input* teks agar panjangnya konsisten dalam *batch*. Lalu ada juga fungsi *Dataset* dari *library datasets* untuk menangani *dataset* dengan efisien. Kode perintah *training* pada Gambar 30 dan Gambar 31.

```

# Mengimpor Trainer dari Hugging Face Transformers
# Trainer API untuk melatih dan mevaluasi model secara efisien
from transformers import Trainer, DataCollatorWithPadding

# Mengimpor Dataset jika Anda ingin mengubah data menjadi
format Hugging Face Dataset (jika diperlukan)
from datasets import Dataset

# Membuat objek Trainer yang akan menangani proses pelatihan
dan evaluasi model

```

```

trainer = Trainer(
    model=model,                                # Model BERT yang sudah
    disiapkan untuk klasifikasi
    args=training_args,                        # Parameter pelatihan yang
    telah ditentukan sebelumnya (TrainingArguments)
    train_dataset=train_dataset,              # Dataset pelatihan (sudah
    dalam format TweetDataset)
    eval_dataset=test_dataset,                # Dataset evaluasi (untuk
    validasi setelah tiap epoch)
    tokenizer=tokenizer,                      # Tokenizer yang digunakan
    untuk menyamakan preprocessing teks
    data_collator=DataCollatorWithPadding(tokenizer)
    # Data collator bertugas melakukan padding otomatis sesuai
    batch input
    # agar tiap input memiliki panjang yang sama dalam satu
    batch
)

```

Gambar 30. Kode Printah *Trainer*

```
# Memulai proses pelatihan model menggunakan Trainer dari
# Proses melatih model pada train_dataset dan mengevaluasinya
pada eval_dataset
trainer.train()
```

Gambar 31. Kode Perintah Penerapan Konfigurasi Pelatihan

4.10 Implementasi Model Klasifikasi

Tahap ini merupakan pengujian dari model klasifikasi dari LSTM dan BERT dengan data uji, Pada LSTM pengujian menggunakan *NumPy* dimana merupakan *library* Python yang digunakan untuk perhitungan numerik dan manipulasi *array*., berikut kode program dari prediksi data uji pada Gambar 32.

```
# Mengimpor library NumPy untuk manipulasi array numerik
import numpy as np

# Menggunakan model untuk memprediksi probabilitas kelas dari
data uji (X_test_pad)
y_pred_prob = model.predict(X_test_pad)

# Mengambil prediksi kelas dengan nilai probabilitas tertinggi
# np.argmax memilih indeks (kelas) dengan nilai tertinggi di
setiap baris (axis=1)
y_pred = np.argmax(y_pred_prob, axis=1)
```

Gambar 32. Kode Program Penerapan Fungsi pada Data Uji LSTM

BERT melakukan prediksi dengan menggunakan *library PyTorch* dan fungsi-fungsi seperti *torch.tensor*, *torch.nn*, *torch.optim*. Kode program penerapan fungsi pada data uji BERT pada Gambar 33.

```

# Mengimpor library PyTorch untuk manipulasi tensor
import torch

# Menggunakan Trainer dari Hugging Face untuk melakukan
prediksi pada test_dataset
# Hasilnya berupa objek PredictionOutput yang berisi prediksi
logits, label, dan metrik evaluasi
preds_output = trainer.predict(test_dataset)

# Mengambil prediksi akhir dengan memilih kelas (label) yang
memiliki nilai logit tertinggi
# - predictions: output mentah dari model (logits)
# - torch.tensor: konversi hasil prediksi ke bentuk tensor
# - torch.argmax: memilih indeks dengan nilai tertinggi di
setiap baris (yaitu label kelas)
preds = torch.argmax(torch.tensor(preds_output.predictions),
axis=1)

```

Gambar 33. Kode Program Penerapan Fungsi pada Data Uji BERT

4.11 Pengujian dan Evaluasi Model Klasifikasi

Pengujian dan evaluasi model dilakukan untuk mengetahui performa LSTM dan BERT. Maka kedua model perlu untuk diuji menggunakan metode *Confusion Matrix*, *Confusion Matrix* merupakan metode yang digunakan untuk mengukur kinerja dari metode klasifikasi, dan merupakan fungsi dari *library sklearn* berikut penerapan metode *Confusion Matrix*.

Tabel 7. Hasil *Confusion Matrix* LSTM

		Predicted Class		
		Negatif	Netral	Positif
Actual Class	Negatif	91	19	20
	Netral	19	91	21
	Positif	20	14	97

Sumber : Hasil Olah Data, 2025

Tabel 8. Hasil *Confusion Matrix* BERT

		Predicted Class		
		Negatif	Netral	Positif
Actual Class	Negatif	118	9	14
	Netral	13	97	16
	Positif	23	20	82

Sumber : Hasil Olah Data, 2025

Setelah mendapatkan *Confusion Matrix*. Penelitian ini juga dapat mengetahui nilai akurasi, *Precision*, *Recall* dan *F1-Score* pada setiap kelas klasifikasi. Nilai *Precision* adalah jumlah data positif yang benar dalam proses klasifikasi dibagi total data jumlah data yang diklasifikasi sebagai positif, Nilai

Recall adalah jumlah data positif yang benar dalam proses klasifikasi dibagi jumlah data positif yang sebenarnya. Sedangkan *F1-Score* merupakan parameter tunggal ukuran keberhasilan *retrieval* yang menggabungkan *Precision* dan *Recall*. Berikut hasil nilai *Precision*, *Recall* dan *F1-Score* dari *Naïve LSTM* dan BERT pada Gambar 34 dan 35.

	precision	recall	f1-score	support
negative	0.70	0.70	0.70	130
neutral	0.73	0.69	0.71	131
positive	0.70	0.74	0.72	131
accuracy			0.71	392
macro avg	0.71	0.71	0.71	392
weighted avg	0.71	0.71	0.71	392

Gambar 34. Hasil Evaluasi Performa LSTM

	precision	recall	f1-score	support
negative	0.77	0.84	0.80	141
neutral	0.77	0.77	0.77	126
positive	0.73	0.66	0.69	125
accuracy			0.76	392
macro avg	0.76	0.75	0.75	392
weighted avg	0.76	0.76	0.76	392

Gambar 35. Hasil Evaluasi Performa BERT

Dari hasil evaluasi performa antara algoritma LSTM dan didapatkan nilai akurasi metode LSTM sebesar 71%, sedikit lebih kecil dari metode BERT dengan nilai akurasi sebesar 76%. Dari perbandingan nilai akurasi dapat disimpulkan bahwa pada penelitian kali ini dengan dataset dan model yang dibangun metode BERT sedikit lebih baik dari metode LSTM. Hasil tersebut dapat bisa terjadi karena jenis dataset, penyeimbangan jumlah dataset *training* serta parameter dan model yang dibangun.

4.12 Visualisasi Data

Pada tahap ini adalah untuk menunjukkan kata apa saja yang sering muncul dengan menggunakan *WordCloud* dan *matplotlib.pyplot*. Berikut adalah kode program untuk menampilkan kata yang sering muncul, dan hasilnya ada pada Gambar 37.

```
# Mengimpor library WordCloud untuk membuat visualisasi awan
kata
from wordcloud import WordCloud

# Mengimpor matplotlib untuk visualisasi
import matplotlib.pyplot as plt
```

```
# Menggabungkan seluruh teks dari kolom 'clean_text' menjadi
satu string besar
# dropna() menghapus nilai kosong, astype(str) memastikan
semua teks bertipe string
text = " ".join(df["hasil_terjemah"].dropna().astype(str))

# Membuat objek WordCloud dengan lebar, tinggi, dan latar
belakang putih
# generate(text) membuat awan kata berdasarkan frekuensi kata
dalam teks
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(text)

# Menentukan ukuran figure (kanvas) untuk menampilkan word
cloud
plt.figure(figsize=(10, 5))

# Menampilkan gambar word cloud menggunakan bilinear
interpolation untuk hasil lebih halus
plt.imshow(wordcloud, interpolation='bilinear')

# Menyembunyikan sumbu x dan y karena tidak diperlukan untuk
visualisasi awan kata
plt.axis("off")

# Menambahkan judul pada grafik
plt.title("Word Cloud dari ", fontsize=16)

# Menampilkan visualisasi di layar
plt.show()
```


[illegible]

Gambar 39. *WordCloud* Sentimen Positif

4.13 Pembahasan

Berdasarkan hasil penelitian pada Analisis Sentimen Pengguna X Terhadap PSSI dan Timnas Indonesia Dalam Perkembangan Sepak Bola Di Indonesia Dengan Algoritma LSTM dan BERT di dapatkan hasil sebagai berikut

1. Hasil dari pemodelaan dengan algoritma LSTM dalam unki analisa sentimen di dapatkan hasil *precision* sebesar 71%, *recall* 71% dan *f1-score* 71% dari 392 data *test* dan didapatkan pula akurasi dalam menguji sentimen adalah 71%, sedangkan pada pemodelan dengan BERT didapatkan nilai *precision* sebesar 76%, *recall* 75%, *f1-score* 75% dari menguji 392 data *test* dan didapatkan nilai akurasi sebesar 76% dengan hasil tersebut dapat disimpulkan pada penelitian ini model BERT lebih baik dibandingkan pada model LSTM, berbanding terbalik dengan penelitian terdahulu yang dilakukan Ezen-can (2020) “*A Comparison of LSTM and BERT for Small Corpus*” berdasarkan hasil penelitian untuk membangun *chatbot* dengan menggunakan dataset yang kecil, didapatkan hasil bahwa pemodelan LSTM memiliki nilai akurasi lebih baik dengan nilai akurasi 70,08%, dibandingkan pemodelan BERT yang memiliki nilai akurasi 67,15%. Hasil ini disebabkan oleh metode pengumpulan data dan proses pengolahan data yang berbeda.
2. Dari penelitian di atas di dapatkan pula hasil dari sentimen pengguna X yang dilakukan menggunakan VADER dan di dapatkan hasil data sentimen positif sebanyak 2185, negatif 652, dan netral sebanyak 1143 data dari total 3983 data. Dari hasil tersebut dapat disimpulkan jika sentimen pengguna X lebih condong kepada positif, dan dari hasil visualisasi didapatkan hasil jika pengguna X memberikan sentimen berfokus kepada pemain, pelatih dan pihak PSSI, dapat disimpulkan senimen pengguna X tergantung pada performa dan kinerja dari Timnas Indonesia dan Juga PSSI.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian Analisis Sentimen Pengguna X Terhadap Kinerja PSSI Dan Timnas Dalam Perkembangan Sepak Bola Di Indonesia dengan menerapkan Algoritma LSTM dan BERT menghasilkan kesimpulan sebagai berikut:

1. Dari penelitian ini didapatkan hasil penelitian berupa akurasi analisa sentimen pada LSTM sebesar 71% dan akurasi pada BERT sebesar 76%, sehingga dapat disimpulkan bahwa penerapan LSTM dan BERT pada penelitian Analisis Sentimen Pengguna X Terhadap Kinerja PSSI Dan Timnas Dalam Perkembangan Sepak Bola Di Indonesia BERT mempunyai tingkat akurasi yang lebih baik.
2. Pada penelitian pelebelaan sentimen yang dilakukan menggunakan VADER didapatkan hasil sentimen yang didominasi positif dengan jumlah mencapai 2185, sedangkan negatif sebesar 652, dan netral berjumlah 1143, dari total data 3981. Lalu dari hasil penelitian dengan LSTM didapatkan pula rata-rata nilai precision, recall dan f1-score memiliki nilai yang sama yaitu sebesar 71%, sedangkan pada algoritma BERT rata-rata nilai precision sebesar 76%, recall 75%, dan f1-score 75%.

5.2 Saran

Berdasarkan hasil penelitian terdapat beberapa saran yang bisa dijadikan acuan untuk meningkatkan hasil penelitian. Diantaranya sebagai berikut:

1. Untuk penelitian berbahasa Indonesia bisa menggunakan *pretrained IndoNLP Word2Vec* untuk LSTM dan *pretrained indoBERT* untuk BERT sehingga tidak perlu menerjemahkan bahasa dengan *library googletrans*.
2. Pada proses pelabelan untuk hasil yang lebih akurat ada baiknya dilakukan secara manual. Karena sebaik-baiknya tool pasti akan ada data yang kurang akurat.
3. Mencoba untuk menggunakan algoritma lain untuk menjadikan perbandingan kualitas antar algoritma.

DAFTAR PUSTAKA

- Arti kata sepak bola - Kamus Besar Bahasa Indonesia (KBBI) Online.* (n.d.). Retrieved July 12, 2025, from [https://kbbi.web.id/sepak bola](https://kbbi.web.id/sepak%20bola)
- Azizah, L. M., Ajipratama, D. B., Putri, N. A. R., & Damarjati, C. (2022). Analisa Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Di Indonesia Pada Twitter Menggunakan Algoritma LSTM La. *JURNAL IPTEKKOM Jurnal Ilmu Pengetahuan & Teknologi Informasi*, 24(2), 161–172. <https://doi.org/10.17933/iptekkom.24.2.2022.161-172>
- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P. (2021). IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 8875–8898. <https://doi.org/10.18653/v1/2021.emnlp-main.699>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Dutch East Indies - Football History.* (n.d.). Retrieved July 12, 2025, from <https://www.rsssf.org/tables/i/indiechamp.html#intro>
- Ezen-Can, A. (2020). *A Comparison of LSTM and BERT for Small Corpus*. 1–12. <http://arxiv.org/abs/2009.05451>
- Farida, Z. W., & Rochmawati, N. (2024). Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di Twitter. *Journal of Informatics and Computer Science (JINACS)*, 5(03), 369–376. <https://doi.org/10.26740/jinacs.v5n03.p369-376>
- George, A. (2022). *Python Text Mining Perform Text Processing, Word Embedding, Text Classification and Machine Translation*.
- Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. In *Synthesis Lectures on Human Language Technologies* (Vol. 10, Issue 1).

<https://doi.org/10.2200/S00762ED1V01Y201703HLT037>

Graves, A. (2013). *Generating Sequences With Recurrent Neural Networks*. 1–43.
<http://arxiv.org/abs/1308.0850>

History of Football - The Origins. (2012). FIFA.Com. <https://www.fifa.com/>

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
<https://doi.org/10.1162/neco.1997.9.8.1735>

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *Association for Computational Linguistics*, 46(2), 119–124.
<https://doi.org/10.3760/cma.j.issn.04124081.2010.02.006>

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for. *Eighth International AAAI Conference on Weblogs and Social Media*, 216–225. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>

Imawan, F., Shiddieq, F., & Roji, F. F. (2025). Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert Analysis of Public Sentiment in X Towards The 12% PPN Increase Plan Using Bert. *Journal of Computer Engineering, System and Science*, 10(1), 136–148.
<https://jurnal.unimed.ac.id/2012/index.php/cess>

Indurkha, N., & Damerau, F. J. (2010). Handbook of Natural Language Processing. *Handbook of Natural Language Processing*.
<https://doi.org/10.1201/9781420085938>

Jo, T. (2024). Text Mining: Concepts, Implementation, and Big Data Challenge: Second Edition. In *Studies in Big Data* (Vol. 45).
<https://doi.org/10.1007/978-3-031-75976-5>

Jurafsky, D., & Martin, J. H. (2021). Speech and language processing: An introduction to speech recognition. *Computational Linguistics and Natural Language Processing. 2nd Edn., Prentice Hall, ISBN, 10(0131873210)*, 794–800.

Khadapi, M., & Maruli Pakpahan, V. (2024). Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi Twitter tentang Pemilu 2024. *JUKI: Jurnal Komputer Dan Informatika*, 6(2), 130–137.

<https://www.ioinformatic.org/index.php/JUKI/article/view/681>

Kulkarni, A., & Shivananda, A. (2019). Natural language processing recipes: Unlocking text data with machine learning and deep learning using python. In *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python*. <https://doi.org/10.1007/978-1-4842-4267-4>

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*.

Maharani, C. A., Warsito, B., & Santoso, R. (2024). Analisis Sentimen Vaksin Covid-19 Pada Twitter Menggunakan Recurrent Neural Network (Rnn) Dengan Algoritma Long Short-Term Memory (Lstm). *Jurnal Gaussian*, 12(3), 403–413. <https://doi.org/10.14710/j.gauss.12.3.403-413>

Manaswi, N. K. (2018). Deep Learning with Applications Using Python. In *Deep Learning with Applications Using Python*. <https://doi.org/10.1007/978-1-4842-3516-4>

Mckinney, W. (2017). *Python for Data Analysis - 2nd edition* (2nd Editio).

Meints, W. (2019). *Deep learning with Microsoft Cognitive Toolkit quick start guide: a practical guide to building neural networks using Microsoft's open source deep learning framework*. <https://search.worldcat.org/title/1100643380>

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends R in Information Retrieval*, 2(2), 1–135. <https://doi.org/10.1561/1500000001>

Russell, M., & Klassen, M. (2019). *Mining the social web: data mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and more*. O'Reilly Media.

Saputra, R. (2025). *Inilah Jenis-Jenis Olahraga Paling Populer di Dunia*. Rri.Co.Id. <https://www.rri.co.id/>

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

Sejarah PSSI. (n.d.). Retrieved July 12, 2025, from <https://www.pssi.org/about/history>

- Sofiana, S. (2024). Konsep BERT pada Natural Language Processing. In *Eureka Media Aksara* (Vol. 3, Issue 1). <http://dx.doi.org/10.1016/j.bpj.2015.06.056%0Ahttps://academic.oup.com/bioinformatics/article-abstract/34/13/2201/4852827%0Ainternal-pdf://semisupervised-3254828305/semisupervised.ppt%0Ahttp://dx.doi.org/10.1016/j.str.2013.02.005%0Ahttp://dx.doi.org/10.10>
- Stevany, R. (2024). *Indonesia Pengguna X atau Twitter Terbanyak Keempat di Dunia*. Rri.Co.Id. <https://www.rri.co.id/>
- Sundermeyer, M., Schlüter, R., & Ney, H. (2012). LSTM neural networks for language modeling. *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012, 1*, 194–197. <https://doi.org/10.21437/interspeech.2012-65>
- Tenney, I., Das, D., & Pavlick, E. (2020). BERT rediscovers the classical NLP pipeline. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 4593–4601. <https://doi.org/10.18653/v1/p19-1452>
- Understanding LSTM Networks -- colah's blog*. (n.d.). Retrieved July 13, 2025, from <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Kaiser, Ł. (2017). Advances in Neural Information Processing Systems. *International Conference on Information and Knowledge Management, Proceedings, Nips*, 4752–4758. <https://doi.org/10.1145/3583780.3615497>
- Vluymans, S. (2019). Learning from imbalanced data. *Studies in Computational Intelligence*, 807(9), 81–110. https://doi.org/10.1007/978-3-030-04663-7_4
- Yoshua Bengio, Ian J. Goodfellow, & Aaron Courville. (2015). Deep Learning. *Clinical Infectious Diseases*, 53(3), 262–268. <https://academic.oup.com/cid/article-lookup/doi/10.1093/cid/cir355>
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). *Recent Trends in Deep Learning Based Natural Language Processing*. <http://veredshwartz.blogspot.sg>.

LAMPIRAN

Lampiran 1. Source Code

a. *Crawling Data*

```
#Twitter Auth Token
twitter_auth_token =
'be5255cab0f1677dbd88e6f9a3b301d8075cfc91'

# Menginstal pustaka Python 'pandas' yang diperlukan untuk
analisis data
!pip install pandas

# Memperbarui daftar paket di sistem agar instalasi
selanjutnya menggunakan versi terbaru
!sudo apt-get update

# Menginstal dependensi yang diperlukan untuk mengatur
repositori Node.js
!sudo apt-get install -y ca-certificates curl gnupg

# Membuat direktori untuk menyimpan keyring (penyimpanan kunci
enkripsi) Node.js
!sudo mkdir -p /etc/apt/keyrings

# Mengunduh dan menyimpan GPG key Node.js ke dalam keyring
untuk memastikan keamanan instalasi
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg

# Menambahkan repositori resmi Node.js versi 20 ke dalam
sources list sistem
!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" |
sudo tee /etc/apt/sources.list.d/nodesource.list

# Memperbarui daftar paket lagi setelah menambahkan repositori
Node.js
!sudo apt-get update

# Menginstal Node.js dari repositori yang telah ditambahkan
!sudo apt-get install nodejs -y
```

```

# Menampilkan versi Node.js yang telah terinstal untuk
verifikasi
!node -v
# Menentukan nama file output untuk menyimpan data hasil crawl
filename = 'data_mentah.csv'

# Menentukan kata kunci pencarian untuk Tweet yang berbahasa
Indonesia,
# mencakup topik "PSSI dan Timnas Indonesia" dari tanggal 27
Juni 2024 hingga 1 Mei 2025
search_keyword = 'PSSI and Timnas Indonesia since:2024-06-27
until:2025-05-01 lang:id'

limit = 4000
# Menjalankan tool tweet-harvest menggunakan NPX untuk
mengambil tweet berdasarkan parameter yang ditentukan:
# -o: nama file output
# -s: kata kunci pencarian
# --tab "LATEST": mengambil tweet terbaru (bukan populer atau
top)
# -l: jumlah maksimum tweet
# --token: token autentikasi Twitter (harus sudah disiapkan
sebelumnya)
!npx -y tweet-harvest@2.6.1 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}

```

b. Terjemah Teks

```

import pandas as pd
df = pd.read_csv('data_mentah.csv')

# Memilih kolom 'full_text' dari DataFrame dan menyimpannya ke
dalam variabel baru
selected_columns = df['full_text']

# Menyimpan kolom 'full_text' ke dalam file CSV baru bernama
'datamentah_baru.csv'
# Parameter index=False digunakan agar indeks baris tidak ikut
disimpan di file CSV
selected_columns.to_csv('datamentah_baru.csv', index=False)

pip install googletrans==4.0.0-rc1

# Mengimpor modul Translator dari pustaka googletrans
from googletrans import Translator

```

```

# Membuat objek Translator untuk melakukan penerjemahan
translator = Translator()

# Fungsi untuk menerjemahkan teks secara aman dari Bahasa
Indonesia ke Bahasa Inggris
def translate_text(text):
    try:
        # Menerjemahkan teks dengan menentukan bahasa sumber
        ('id') dan tujuan ('en')
        return translator.translate(str(text), src='id',
dest='en').text
    except:
        # Jika terjadi error (misalnya koneksi atau teks
        kosong), kembalikan teks aslinya
        return text

# Menerapkan fungsi translate_text ke kolom 'full_text' dalam
DataFrame
# Hasil terjemahan disimpan ke kolom baru bernama
'hasil_terjemah'
df['hasil_terjemah'] = df['full_text'].apply(translate_text)

import pandas as pd
df = pd.read_csv('datamentah_baru.csv')
df

```

c. Pelabelan Data

```

# Mengimpor pustaka yang diperlukan
import nltk # Pustaka untuk pemrosesan bahasa alami
from nltk.sentiment import SentimentIntensityAnalyzer # Kelas
untuk analisis sentimen
import matplotlib.pyplot as plt # Pustaka untuk visualisasi
grafik

# Mengunduh lexicon VADER yang diperlukan
nltk.download('vader_lexicon')

# Inisialisasi objek SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

# Fungsi untuk menentukan sentimen berdasarkan skor komposit
def get_sentiment(text):
    score = sia.polarity_scores(text) # Menghitung skor
    sentimen
    compound = score['compound'] # Mengambil skor komposit
    if compound >= 0.05:
        return 'positive' # Sentimen positif

```



```

elif compound <= -0.05:
    return 'negative' # Sentimen negatif
else:
    return 'neutral' # Sentimen netral

# Menerapkan fungsi get_sentiment pada kolom 'hasil_terjemah'
DataFrame df
df['sentiment'] = df['hasil_terjemah'].apply(get_sentiment)

# Visualisasi distribusi sentimen
df['sentiment'].value_counts().plot(kind='bar',
color=['green', 'red', 'gray'])
plt.title('Sentiment Distribution (VADER)') # Judul grafik
plt.xlabel('Sentiment') # Label sumbu x
plt.ylabel('Tweet Count') # Label sumbu y
plt.show() # Menampilkan grafik
# Menampilkan jumlah data berdasarkan kategori sentimen
print(df['sentiment'].value_counts())

```

d. Penanganan *Imbalance Data*

```

# Tujuannya untuk menyeimbangkan jumlah data antar kelas
min_count = df['sentiment'].value_counts().min()

# Menyeimbangkan dataset dengan cara mengambil sampel acak
dari setiap kelas sebanyak 'min_count'
# Sehingga semua kelas memiliki jumlah data yang sama
df_balanced = df.groupby('sentiment').apply(lambda x:
x.sample(min_count)).reset_index(drop=True)

# Mengimpor library matplotlib untuk visualisasi dasar
import matplotlib.pyplot as plt

# Mengimpor seaborn untuk visualisasi yang lebih menarik dan
interaktif
import seaborn as sns

# Menampilkan jumlah data pada setiap kelas sentimen dalam
bentuk angka (output ke konsol)
print(df_balanced['sentiment'].value_counts())

# Membuat visualisasi jumlah tweet per kelas sentimen
menggunakan countplot dari seaborn
sns.countplot(x='sentiment', data=df_balanced)

# Memberikan judul pada grafik
plt.title('Distribusi Sentimen Tweet')

```

```
# Menampilkan grafik ke layar
plt.show()
```

e. Preprocessing Teks

```
import re
import nltk
import pandas as pd
from nltk.corpus import stopwords

# Unduh stopwords bahasa Inggris (sekali saja)
nltk.download('stopwords')

# Set stopwords bahasa Inggris
stop_words = set(stopwords.words('english'))

# Fungsi: Pembersihan teks (lowercase, hapus URL, mention,
# hashtag, simbol)
def clean_text(text):
    text = text.lower()
    text = re.sub(r"http\S+", "", text)           # Hapus URL
    text = re.sub(r"@w+", "", text)               # Hapus mention
    (@user)
    text = re.sub(r"#w+", "", text)               # Hapus hashtag
    (#tag)
    text = re.sub(r"[^a-zA-Z\s]", "", text)       # Hapus
    karakter non-alfabet
    text = re.sub(r"\s+", " ", text).strip()      # Hapus spasi
    berlebih
    return text

# Fungsi: Stopword removal
def remove_stopwords_en(text):
    words = text.split()
    filtered = [word for word in words if word not in
stop_words]
    return " ".join(filtered)

# 1. Bersihkan teks
df['clean_text'] = df['hasil_terjemah'].apply(clean_text)

# 2. Hapus stopwords
df['clean_text'] = df['clean_text'].apply(remove_stopwords_en)
```

f. Encode Label

```
pip install transformers datasets scikit-learn pandas torch
numpy
```

```
# Mengimpor LabelEncoder dari scikit-learn untuk mengubah
label kategorikal menjadi angka
from sklearn.preprocessing import LabelEncoder

# Membuat objek LabelEncoder
label_encoder = LabelEncoder()

# Mengubah nilai pada kolom 'sentiment' menjadi angka, dan
menyimpannya ke kolom baru 'label_encoded'
# Contoh: 'positif' → 2, 'netral' → 1, 'negatif' → 0 (urutan
tergantung isi data)
df['label_encoded'] =
label_encoder.fit_transform(df['sentiment'])

# Menyimpan jumlah kelas unik dalam variabel num_classes
(berguna untuk keperluan pelatihan model)
num_classes = len(label_encoder.classes_)
```

```
df[['sentiment', 'label_encoded']].drop_duplicates()
```

g. Split Data

```
# Mengimpor fungsi train_test_split dari scikit-learn untuk
membagi dataset menjadi data latih dan data uji
from sklearn.model_selection import train_test_split

# Membagi data menjadi data latih (80%) dan data uji (20%)
dengan mempertahankan proporsi label (stratifikasi)
train_texts, test_texts, train_labels, test_labels =
train_test_split(
    df['hasil_terjemah'],      # Fitur input berupa teks yang
    df['label_encoded'],       # Label target dalam
    bentuk numerik
    test_size=0.2,             # 20% data digunakan
    sebagai data uji
    random_state=42,           # Seed untuk memastikan
    hasil pembagian data tetap sama jika dijalankan ulang
    stratify=df['label_encoded'] # Memastikan distribusi
    label di data latih dan data uji tetap proporsional
)
```

h. Tokenisasi LSTM

```
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import
pad_sequences
import numpy as np
```

```

tokenizer = Tokenizer(num_words=10000, oov_token="<OOV>") # 1.
Tokenizer dengan num_words besar dan oov_token
tokenizer.fit_on_texts(train_texts) # X_train adalah list of
text (teks mentah)

X_train_seq = tokenizer.texts_to_sequences(train_texts) #
Tokenisasi
oov_count = sum([seq.count(1) for seq in X_train_seq]) # 2.
Hitung jumlah token <OOV> (biasanya token 1)

lengths = [len(seq) for seq in X_train_seq] # 3. Analisis
panjang input untuk menentukan max_len
print("Panjang rata-rata kalimat:", np.mean(lengths))
print("Panjang 90% kalimat:", np.percentile(lengths, 90))

max_len = int(np.percentile(lengths, 90)) # Gunakan panjang
ke-90 persentil sebagai max_len
print(f"max_len yang disarankan: {max_len}")

X_train_pad = pad_sequences(X_train_seq, maxlen=max_len,
padding='post', truncating='post') # Padding sequence

```

i. Encode One-Hot Encoding Label

```

# Mengimpor fungsi to_categorical dari TensorFlow Keras
# Fungsi ini digunakan untuk mengubah label numerik menjadi
format one-hot encoding
from tensorflow.keras.utils import to_categorical

# Mengubah label pelatihan (train_labels) menjadi bentuk one-
hot encoding
# Misalnya label 2 akan menjadi [0, 0, 1] jika ada 3 kelas
y_train_cat = to_categorical(train_labels,
num_classes=num_classes)

# Mengubah label pengujian (test_labels) menjadi bentuk one-
hot encoding
y_test_cat = to_categorical(test_labels,
num_classes=num_classes)

```

j. Tokenisasi BERT

```

# Mengimpor tokenizer dari library Hugging Face Transformers
# BertTokenizer digunakan untuk mempersiapkan teks agar bisa
diproses oleh model BERT
from transformers import BertTokenizer

# Memuat pre-trained tokenizer dari model 'bert-base-uncased'
(huruf kecil semua, bahasa Inggris)

```

```

tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')

# Tokenisasi teks data latih:
# - Mengonversi teks menjadi token/token ID
# - Menambahkan padding agar panjang input seragam
# - Melakukan pemotongan (truncation) jika melebihi panjang
maksimal BERT (biasanya 512 token)
train_encodings = tokenizer(list(train_texts),
truncation=True, padding=True)

# Tokenisasi teks data uji dengan cara yang sama
test_encodings = tokenizer(list(test_texts), truncation=True,
padding=True)

```

k. Menyiapkan Dataset Tokenisasi BERT Untuk Pelatihan BERT

```

# Mengimpor library PyTorch
import torch

# Membuat class dataset kustom untuk tweet, yang sesuai dengan
format yang dibutuhkan oleh model BERT di PyTorch
class TweetDataset(torch.utils.data.Dataset):
    # Konstruktor (__init__) untuk menyimpan hasil tokenisasi
(encodings) dan label (labels)
    def __init__(self, encodings, labels):
        self.encodings = encodings # dictionary: input_ids,
attention_mask, dll
        self.labels = labels # list label untuk setiap
sampel

    # Method __getitem__ akan mengambil satu item berdasarkan
indeks (idx)
    def __getitem__(self, idx):
        # Mengambil input sesuai index dan mengubahnya ke
tensor PyTorch
        item = {k: torch.tensor(v[idx]) for k, v in
self.encodings.items()}
        # Menambahkan label ke dalam item (juga sebagai
tensor)
        item['labels'] = torch.tensor(self.labels[idx])
        return item # item akan berisi input_ids,
attention_mask, dan labels dalam bentuk tensor

    # Method __len__ untuk mengembalikan panjang dataset
(jumlah sampel)
    def __len__(self):
        return len(self.labels)

```

```
# Membuat dataset PyTorch dari data latih (train)
train_dataset = TweetDataset(train_encodings,
list(train_labels))

# Membuat dataset PyTorch dari data uji (test)
test_dataset = TweetDataset(test_encodings, list(test_labels))
```

```
sample = train_dataset[0]
print(sample.keys()) # dict_keys(['input_ids',
'attention_mask', 'labels'])
print(sample['input_ids'][:10]) # 10 token pertama
print(sample['labels']) # label kelas
```

1. Membangun Model LSTM

```
# Mengimpor modul untuk membangun model neural network
menggunakan Keras
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Embedding, LSTM, Dropout,
Dense

# Membangun model sekuensial (berlapis satu per satu)
model = Sequential([
    # Layer embedding untuk mengubah token integer menjadi
representasi vektor berdimensi 128
    # input_dim = jumlah kata unik (vocab size), output_dim =
dimensi vektor, input_length = panjang input tetap
    Embedding(input_dim=10000, output_dim=128,
input_length=max_len),

    # Layer LSTM untuk menangkap informasi urutan dalam teks
(recurrent layer)
    # return_sequences=False berarti hanya mengambil output
terakhir, cocok untuk klasifikasi
    LSTM(128, return_sequences=False),

    # Dropout layer untuk mengurangi overfitting dengan
membuang 50% neuron secara acak saat training
    Dropout(0.5),

    # Dense layer (fully connected) dengan 64 neuron dan
aktivasi ReLU
    Dense(64, activation='relu'),

    # Output layer dengan neuron sebanyak jumlah kelas dan
aktivasi softmax untuk klasifikasi multi-kelas
    Dense(num_classes, activation='softmax')
```

```

])

# Menyusun model dengan:
# - loss function: categorical_crossentropy (karena one-hot
encoding digunakan)
# - optimizer: Adam (algoritma optimisasi populer)
# - metrics: akurasi untuk evaluasi kinerja
model.compile(loss='categorical_crossentropy',
optimizer='adam', metrics=['accuracy'])

# Menampilkan ringkasan arsitektur model
model.summary()

```

m. Membangun Model BERT

```

# Mengimpor model BERT untuk tugas klasifikasi dari library
Transformers (Hugging Face)
from transformers import BertForSequenceClassification

# Memuat model BERT pre-trained ('bert-base-uncased') yang
disesuaikan untuk klasifikasi teks
# - num_labels menunjukkan jumlah kelas yang akan diprediksi
oleh model
# - Model ini secara otomatis menambahkan layer klasifikasi di
atas arsitektur BERT
model = BertForSequenceClassification.from_pretrained(
    'bert-base-uncased',                # Model BERT
    pre-trained dengan lowercase text
    num_labels=len(label_encoder.classes_) # Menentukan
jumlah kelas target (misalnya: positif, netral, negatif)
)

```

n. Training Data LSTM

```

history = model.fit(
    X_train_pad,      # Data teks yang sudah diproses dan
dipadatkan (padding)
    y_train_cat,      # Label dalam bentuk one-hot encoding
    validation_split=0.2, # 20% data training dipakai sebagai
validasi
    epochs=20,        # Model dilatih sebanyak 20 kali
iterasi terhadap seluruh data
    batch_size=32      # Jumlah sampel per batch saat
training
)

```

o. TrainingArgs Pada BERT

```

# Mengimpor TrainingArguments dari library Hugging Face
Transformers

```

```

# TrainingArguments digunakan untuk mengatur parameter
pelatihan model
from transformers import TrainingArguments

# Menentukan konfigurasi/parameter untuk proses pelatihan
model
training_args = TrainingArguments(
    output_dir='./results',          # Direktori untuk
    # Menyimpan model yang telah dilatih
    num_train_epochs=3,              # Jumlah epoch
    # (putaran pelatihan penuh terhadap seluruh data)
    per_device_train_batch_size=8,    # Ukuran batch
    # saat pelatihan untuk setiap perangkat (misalnya per GPU)
    per_device_eval_batch_size=8,     # Ukuran batch
    # saat evaluasi untuk setiap perangkat
    eval_strategy="epoch",            # Melakukan
    # evaluasi setiap selesai satu epoch
    save_strategy="epoch",            # Menyimpan model
    # setiap selesai satu epoch
    logging_dir='./logs',             # Direktori untuk
    # menyimpan log pelatihan
    load_best_model_at_end=True        # Secara otomatis
    # memuat model terbaik (berdasarkan evaluasi) setelah training
    # selesai
)

```

p. Membuat *Trainer* Untuk BERT

```

# Mengimpor Trainer dari Hugging Face Transformers
# Trainer adalah API tingkat tinggi untuk melatih dan
# mengevaluasi model secara efisien
from transformers import Trainer, DataCollatorWithPadding

# Mengimpor Dataset jika Anda ingin mengubah data menjadi
# format Hugging Face Dataset (jika diperlukan)
from datasets import Dataset

# Membuat objek Trainer yang akan menangani proses pelatihan
# dan evaluasi model
trainer = Trainer(
    model=model,                    # Model BERT yang sudah
    # disiapkan untuk klasifikasi
    args=training_args,             # Parameter pelatihan yang
    # telah ditentukan sebelumnya (TrainingArguments)
    train_dataset=train_dataset,    # Dataset pelatihan (sudah
    # dalam format TweetDataset)
    eval_dataset=test_dataset,      # Dataset evaluasi (untuk
    # validasi setelah tiap epoch)
)

```



```

        tokenizer=tokenizer,          # Tokenizer yang digunakan
        untuk menyamakan preprocessing teks
        data_collator=DataCollatorWithPadding(tokenizer)
        # Data collator bertugas melakukan padding otomatis sesuai
        batch input
        # agar tiap input memiliki panjang yang sama dalam satu
        batch
    )

```

q. **Training BERT**

```

# Memulai proses pelatihan model menggunakan Trainer dari
Hugging Face
# Proses ini akan melatih model pada train_dataset dan
mengevaluasinya pada eval_dataset
trainer.train()

```

r. **Pengujian Model LSTM**

```

# Mengimpor library NumPy untuk manipulasi array numerik
import numpy as np

# Menggunakan model untuk memprediksi probabilitas kelas dari
data uji (X_test_pad)
y_pred_prob = model.predict(X_test_pad)

# Mengambil prediksi kelas dengan nilai probabilitas tertinggi
# np.argmax memilih indeks (kelas) dengan nilai tertinggi di
setiap baris (axis=1)
y_pred = np.argmax(y_pred_prob, axis=1)

```

s. **Pengujian Model BERT**

```

# Mengimpor library PyTorch untuk manipulasi tensor
import torch

# Menggunakan Trainer dari Hugging Face untuk melakukan
prediksi pada test_dataset
# Hasilnya berupa objek PredictionOutput yang berisi prediksi
logits, label, dan metrik evaluasi

preds_output = trainer.predict(test_dataset)

# Mengambil prediksi akhir dengan memilih kelas (label) yang
memiliki nilai logit tertinggi
# - predictions: output mentah dari model (logits)
# - torch.tensor: konversi hasil prediksi ke bentuk tensor
# - torch.argmax: memilih indeks dengan nilai tertinggi di
setiap baris (yaitu label kelas)

```

```
preds = torch.argmax(torch.tensor(preds_output.predictions),
axis=1)
```

t. Evaluasi Hasil dan *Confusion Matrix* LSTM

```
# Menampilkan laporan klasifikasi yang mencakup metrik
precision, recall, f1-score, dan support
print("\nClassification Report:")
print(classification_report(test_labels, y_pred,
target_names=label_encoder.classes_))

# Membuat confusion matrix untuk melihat jumlah prediksi benar
dan salah untuk setiap kelas
cm = confusion_matrix(test_labels, y_pred)

# Menampilkan confusion matrix dalam bentuk visual (plot)
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=label_encoder.classes_)
disp.plot(cmap='Blues') # cmap='Blues' memberikan warna biru
pada grafik
plt.title("Confusion Matrix - LSTM Sentiment Analysis") #
Judul grafik
plt.show() # Menampilkan grafik
```

u. Evaluasi Hasil dan *Confusion Matrix* BERT

```
# Evaluation
# Menampilkan laporan klasifikasi yang berisi metrik evaluasi
seperti:
# precision, recall, f1-score, dan support untuk masing-masing
kelas sentimen
print("\nClassification Report:")
print(classification_report(test_labels, preds,
target_names=label_encoder.classes_))

# Confusion Matrix
# Membuat confusion matrix untuk membandingkan prediksi model
vs label sebenarnya

cm = confusion_matrix(test_labels, preds)

# Menampilkan confusion matrix dalam bentuk visual agar lebih
mudah dianalisis

disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=label_encoder.classes_)
disp.plot(cmap="Blues") # Menggunakan colormap biru untuk
tampilan grafik
```

```
plt.title("Confusion Matrix - Sentiment Analysis") # Judul
dari confusion matrix
plt.show() # Menampilkan grafik
```

v. Visualisasi WordCloud

```
# Mengimpor library WordCloud untuk membuat visualisasi awan
kata
from wordcloud import WordCloud

# Mengimpor matplotlib untuk visualisasi
import matplotlib.pyplot as plt

# Menggabungkan seluruh teks dari kolom 'clean_text' menjadi
satu string besar
# dropna() menghapus nilai kosong, astype(str) memastikan
semua teks bertipe string
text = " ".join(df["hasil_terjemah"].dropna().astype(str))

# Membuat objek WordCloud dengan lebar, tinggi, dan latar
belakang putih
# generate(text) membuat awan kata berdasarkan frekuensi kata
dalam teks
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(text)

# Menentukan ukuran figure (kanvas) untuk menampilkan word
cloud
plt.figure(figsize=(10, 5))

# Menampilkan gambar word cloud menggunakan bilinear
interpolation untuk hasil lebih halus
plt.imshow(wordcloud, interpolation='bilinear')

# Menyembunyikan sumbu x dan y karena tidak diperlukan untuk
visualisasi awan kata
plt.axis("off")

# Menambahkan judul pada grafik
plt.title("Word Cloud dari ", fontsize=16)

# Menampilkan visualisasi di layar
plt.show()
```

w. WordCloud Sentimen Negatif

```
# 1. Filter baris dengan kolom 'sentiment' == 'negative'
df_negatif = df[df['sentiment'] == 'negative']
```

```

# 2. Gabungkan semua teks bersih (clean_text), buang NaN,
pastikan string
text_negatif = "
".join(df_negatif["clean_text"].dropna().astype(str))
# 3. Buat objek WordCloud untuk teks negatif
wordcloud_negatif = WordCloud(
    width=800,
    height=400,
    background_color='white',          # Latar putih agar kata-
kata mudah terlihat
    colormap='Reds',                   # Skema warna merah untuk
nuansa negatif :contentReference[oaicite:1]{index=1}
    max_words=200,                     # Batasi jumlah kata agar
visual tetap fokus
    stopwords=None                     # Bisa tambahkan daftar
stopwords jika diperlukan
).generate(text_negatif)

# 4. Visualisasi WordCloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud_negatif, interpolation='bilinear') #
Rendering halus
plt.axis("off")                                #
Hilangkan axis
plt.title("Word Cloud Sentimen Negatif", fontsize=16)  #
Judul yang informatif
plt.show()

```

x. WordCloud Sentimen Positif

```

# 1. Filter tweet berlabel 'positive'
df_positif = df[df['sentiment'] == 'positive']

# 2. Gabungkan semua teks bersih (clean_text), membuang nilai
NaN, pastikan bertipe string
text_positif = "
".join(df_positif["clean_text"].dropna().astype(str))

# 3. Buat WordCloud untuk sentimen positif
wordcloud_positif = WordCloud(
    width=800,
    height=400,
    background_color='white',          # Latar putih untuk
kontras tinggi
    colormap='Greens'                  # Warna hijau umum untuk
visualisasi positif :contentReference[oaicite:1]{index=1}
).generate(text_positif)

```

```
# 4. Visualisasi WordCloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud_positif, interpolation='bilinear') #
Interpolasi halus untuk tampilan lebih elegan
plt.axis("off") #
Hilangkan sumbu agar fokus ke WordCloud
plt.title("Word Cloud Sentimen Positif", fontsize=16) #
Tambahkan judul yang jelas
plt.show()
```

Lampiran 2. Hasil Terjemah Kata

full_text	hasil_terjemah
Timnas Indonesia Berencana Hadapi Malaysia dan Lebanon pada September 2025 https://t.co/cvL3OdcUNy	The Indonesian national team plans to face Malaysia and Lebanon in September 2025 https://t.co/cvL3OdcunY
Ole Romeny Blak-blakan Mau Bela Timnas Indonesia Berkat Fans Garuda https://t.co/81kagJYTbi	Ole Romeny, I want to defend the Indonesian national team thanks to Garuda fans https://t.co/81Kagiytbi
Asisten Pelatih Timnas Indonesia Denny Landzaat Dibidik Jadi Pelatih Willem II https://t.co/2xhXWw6z1H	Indonesian national team assistant coach Denny Landzaat is targeted to be Willem II coach https://t.co/2xhxww6z1h
Timnas putri Indonesia umumkan 26 nama untuk FIFA Matchday kontra Yordania dan Bangladesh. Jadi persiapan menuju ASEAN Women's Championship. https://t.co/Eau9ZDvhrb	Indonesian women's national team announced 26 names for FIFA Matchday against Jordanian and Bangladesh. So preparation for the ASEAN Women's Championship. https://t.co/eau9zdvhrb
Pratama Arhan Mulai Kehilangan Tempat di Timnas Indonesia Ayah Azizah Salsha Bilang Begini https://t.co/Nvne5h0Vxe	Pratama Arhan began to lose place in the Indonesian national team Azizah Salsha's father said this https://t.co/nvne5h0vxe
5 Pemain Naturalisasi Tambahan yang Dijagokan Dipanggil Patrick Kluivert untuk Laga Timnas Indonesia vs China dan Jepang https://t.co/0imZwJyOBl	5 Additional Naturalization Players Championed Patrick Kluivert for the Indonesian National Team Match vs China and Japan https://t.co/0imzwjyobl
@idextratime Uda kayak pertahanan timnas Indonesia 10 tahun yang lalu kalo ketemu lawan tim Arab	@Idextratime has been like the Indonesian national team defense 10 years ago if you meet against the Arab team
Detail best offer #betis1 Kamis 1 Mei 2025 Timnas Indonesia home 2012/14 Size L 76 x 54 cm Brand new with tag NNS remake (-) 1 mikro hole nyaris tak terlihat di punggung kiri Mulus ganteng pisan ieu BIN: 1.500k or best offer! https://t.co/4v4Pq8W8Yi	Details Best Offer #Betis1 Thursday 1 May 2025 Indonesian National Team Home 2012/14 Size L 76 x 54 cm Brand New with Tag NNS Remake (-) 1 Micro Hole is barely visible on the left back smoothly handsome ieu bin: 1,500k or best offer! https://t.co/4v4pq8w8yi
Alex Pastoor Ngamuk Agenda Lawan Timnas Indonesia vs Rusia Jadi Polemik Baru Denis Manturov Angkat Bicara https://t.co/bUH5lOmeHT	Alex Pastoor Removing the agenda of the Indonesian National Team vs Russia so the new polemic Denis Manturov spoke up https://t.co/buh5lomeht
Detail best offer #betis11 Kamis 1 Mei 2025 Timnas Indonesia official parka jacket 2024 Size M 71 x 55 cm Brand new with tag Mulus tak ada minus. Note: tidak dijual komersil diproduksi hanya untuk dipakai pemain. LIMITED dan RARE! BIN: 850k or best offer! https://t.co/4RbcR0C1dJ	Detail Best Offer #Betis11 Thursday 1 May 2025 Indonesian National Team Official Parka Jacket 2024 Size M 71 x 55 cm Brand New with smooth tags None Minus. Note: Commercial sold is produced only for players to use. Limited and Rare! Bin: 850k or best offer! https://t.co/4rbcr0c1dj

@Jerseyforum Udah dipake asisten timnas Indonesia https://t.co/Sl455XtPYp	@Jerseyforum has been used by the Indonesian national team assistant https://t.co/sl455XTPYP
Timnas Indonesia Jadi Tuan Rumah Babak Keempat Kualifikasi Piala Dunia 2026 Zona Asia jika Penuhi Syarat Ini? https://t.co/qbge6sezwX	The Indonesian national team is hosting the fourth round of the 2026 Asian World Cup qualifiers if they meet this condition? https://t.co/qbge6sezwX
Patrick Kluivert Panggil Ciro Alves ke Timnas Indonesia jika Resmi Jadi WNI? https://t.co/hPoqNVS7mA	Patrick Kluivert Calls Ciro Alves to the Indonesian National Team if officially become a citizen? https://t.co/hpoqnvs7ma
[Hemat hingga 58%] Rasakan kenyamanan dan gaya dengan Baju Jersey Sepak Bola Timnas Indonesia 2025. https://t.co/XxP7XXlMcM	[Saving up to 58%] Feel comfort and style with Indonesian national team soccer jersey 2025. https://t.co/xxp7xxlmcm
Tertarik mengetahui gaji pemain Timnas U-20 Indonesia? Salah satu pemain andalan Iqbal Gwijangge diketahui menerima gaji yang cukup fantastis. Cek artikel lengkapnya di sini â€” https://t.co/KLPdOMneFJ #GajiPemainTimnasU20 #IqbalGwijangge #KarierSepakBola #Kekitaan	Interested in knowing the salaries of Indonesian U-20 national team players?One of the mainstay players of Iqbal Gwijangge is known to receive a fairly fantastic salary.Check the full article here â€” https://t.co/klpDomnefj #gajipemaintimnasu20 #iqbalgwijangge #kariersjkbola #kekuananan
Anak pelatih Timnas Indonesia Shane Kluivert dengan trophy UYL. ðŸ†³ðŸ†± https://t.co/4grAwu63ow	Indonesian National Team Coach Son Shane Kluivert with Trophy Uyl.ðŸ†³ðŸ†± https://t.co/4grawu63ow
Apa yg kalian cari? Ngebon pemain mencekik peluang pemain lokal. @erickthohir @PSSI https://t.co/OdCcVvx3Ke	What are you looking for?Ngebon players strangle the opportunity for local players.@erickthohir @ssi https://t.co/odccvx3ke
Pratama arhan punggawa skuad garuda timnas indonesia telah menyelesaikan pertandingan bersama klubnya bangkok united di lanjutan liga 1 Thailand pekan ke 30& ini rapornya. https://t.co/O2zlwoqKqT	Pratama Arhan, retainer of the Garuda Indonesian National Team squad, has finished the match with his club Bangkok United in the continued League 1 Thailand Week 30 & Amp; https://t.co/o2ziwoqkqt
Cek Jersey Bola Timnas Indonesia 2025 Dryfit Dewasa dengan harga Rp117.520. Dapatkan di Shopee sekarang! https://t.co/tXVi4Cvzkz https://t.co/rb8GwTGXVA	Check the Indonesian National Team Football Jersey 2025 Dryfit Adult at a price of Rp117,520.Get it in Shopee now! https://t.co/txvi4cvzkz https://t.co/rb8gwtgxva
Cek Jersey Timnas Indonesia 2025 Grade Ori Dryfit Dewasa dengan harga Rp117.520. Dapatkan di Shopee sekarang! https://t.co/OWRvflOdFZ https://t.co/mvrWu981em	Check the Indonesian National Team Jersey 2025 Ori Dryfit Adult Grade at a price of Rp117,520.Get it in Shopee now! https://t.co/owrvflodfz https://t.co/mvrwu981em
Erick Thohir Menolak Gagasan Jadi Ketua Umum PSSI Seumur Hidup https://t.co/KFqTXhjNPK	Erick Thohir rejected the idea of becoming the General Chair of PSSI for life https://t.co/kfqtxhjnpk

@DhankzNazie @ErikaUtami18 @RPurnama68781 @kompascom @UGMYogyakarta Di Indonesia itu ga ada yg mungkin bro selagi punya uang semua bisa dikondisikan.. Ingat dokter gadungan yg jadi tim medis buat timnas PSSI??.. Kurang absurd gimana coba itu	@Dhankznazie @erikautami18 @rpurnama68781 @kompascom @ugmyogyakarta in Indonesia there is nothing possible bro while having money all can be conditioned .. remember the fake doctor who became the medical team for the PSSI national team ?? .. less absurd h
Ada 6 negara yang tampil di ronde ke-4 Kualifikasi Piala Dunia 2026 zona Asia dan akan memperebutkan 2 tiket lolos otomatis. Uni Emirat Arab dan Irak difavoritkan jadi tuan rumah! Bagaimana dengan Timnas Indonesia? â³ https://t.co/WtwxzslsQl	There are 6 countries that appear in the 4th round of the 2026 Asian World Cup qualifiers and will compete for 2 automatic qualify tickets.The United Arab Emirates and Iraq are favored to be the host!What about the Indonesian national team?â³ https://t.co/wtwxzslsqi
Asisten Pelatih Timnas Indonesia Denny Landzaat Dibidik Jadi Pelatih Willem II https://t.co/2xhXWw76Rf	Indonesian national team assistant coach Denny Landzaat is targeted to be Willem II coach https://t.co/2xhxww76rf
Pratama Arhan Mulai Kehilangan Tempat di Timnas Indonesia Ayah Azizah Salsha Bilang Begini https://t.co/Nvne5h0nHG	Pratama Arhan began to lose its place in the Indonesian national team Azizah Salsha's father said this https://t.co/nvne5h0nhg
Gaji Robi Darwis gelandang bertahan Timnas Indonesia bikin penasaran? Intip berapa pendapatannya sebagai pemain muda yang bersinar https://t.co/IOs0KGr5lL #RobiDarwis #TimnasIndonesia #GajiPemainBola #SepakBolaIndonesia #Gaji2025 #InfoGaji #PemainMuda	Robi Darwis's Salary Defensing midfielder Indonesian national team is curious?Take a peek at how much income as a shining young player https://t.co/ios0kgr5ll #robidarwis #timnasindonesia #gajipemainbola #sepakbolaindonesia #salary2025 #infogaji #pemainmuda
Gelandang bertahan Timnas Indonesia Rivaldo Pakpahan jadi sorotan! Tapi... berapa ya gajinya sekarang? Cek detailnya di artikel ini: https://t.co/GNINXtQNiA #RivaldoPakpahan #GajiPemainBola #GajiTimnasIndonesia #SepakBolaIndonesia #GajiAtlet2025 #BeritaBola #InfoGaji	Indonesian national team defensive midfielder Rivaldo Pakpahan is in the spotlight!But ... how much is the salary now?Check the details in this article: https://t.co/gognlnxtqnia #rivaldopakpahan #gajipemainbola #gajitimnasindonesia #sepakbolaindonesia #Salarytlet2025 #beritabola #infogaji
Joey Pelupessy Sebut 1 Keistimewaan Skuad Timnas Indonesia Era Patrick Kluivert https://t.co/bdBX4oA3gO	Joey Pelupessy Called 1 Features of the Indonesian National Team Squad Patrick Kluivert https://t.co/bdbx4oa3Go
Sebanyak 4 pemain PSS Sleman sempat absen saat melawan Persib Bandung pada pekan ke-30 Liga 1 2024-2025. Kini pada pemain tersebutu sudah kembali berlatih bersama anak asuh Pieter Huistra lainnya. https://t.co/NJpJPPFebj	A total of 4 PSS Sleman players were absent against Persib Bandung in the 30th week of League 1 2024-2025.Now the player has returned to training with other Pieter Huistra foster children.https://t.co/njppjppfebj
Sebanyak 4 pemain PSS Sleman sempat absen saat melawan Persib Bandung pada	A total of 4 PSS Sleman players were absent against Persib Bandung in the 30th

pekan ke-30 Liga 1 2024-2025. Kini pada pemain tersebut sudah kembali berlatih bersama anak asuh Pieter Huistra lainnya. https://t.co/nWArE9Pw9V	week of League 1 2024-2025. Now the player has returned to training with other Pieter Huistra foster children. https://t.co/nware9pw9v
Ole Romeny Mengenang Golnya bersama Timnas Indonesia Striker timnas Indonesia Ole Romeny mengenang momen spesialnya bersama tim Garuda saat dirinya mencatatkan debut penampilan dan debut gol kala menghadapi Australia di Stadion Sydney Maret. Dikutip dari laman resmi Kita https://t.co/V9cczeXmNq	Ole Romeny recalled his goal with the Indonesian national team striker Indonesian national team Ole Romeny recalled his special moment with the Garuda team when he made his debut appearance and goal debut when facing Australia at the Sydney Stadium March. Quoted from our official page https://t.co/v9cczexmnq
@KetepAcid Saya tidak memiliki informasi terkini tentang rumor transfer Jay Idzes hingga April 2025. Namun berdasarkan data sebelumnya Jay Idzes pemain Venezia dan timnas Indonesia dikaitkan dengan klub seperti Inter Milan dan Bologna. Penampilannya di Se	@Ketepacid I do not have the latest information about the rumors of Jay Idzes' transfer until April 2025. But based on previous data Jay Idzes Venezia players and Indonesian national teams are associated with clubs such as Inter Milan and Bologna. His appe
Sayangnya timnas Indonesia buat next FIFA matchday masih tanpa Shalika https://t.co/L8XXx8zFke	Unfortunately the Indonesian national team for Next FIFA Matchday is still without Shalika https://t.co/l8xxx8zfke
AFC Ubah Format Babak 4 Kualifikasi Piala Dunia 2026? Timnas Indonesia Bakal Diuntungkan https://t.co/9irf4ZdILQ	AFC changed the format of Round 4 of the 2026 World Cup Qualifiers? Indonesian national team will benefit https://t.co/9irf4zdilq
Nilai Pasar Pemain Timnas Indonesia Ini Paling Besar Tertinggi Rp 128 Miliar https://t.co/9Xzxly5kvY	The market value of Indonesian national team players is the highest of Rp 128 billion https://t.co/9xxziy5kvY
@aingriwehuy Gpp di timur Tengah juga kan orang Indonesia banyak yg tinggal di sana bisa tetep militan mendukung timnas	@aingriwehuy GPP in the Middle East also many Indonesians who live there can still be militant to support the national team
Cek Jersey Timnas Indonesia Grade Ori dengan harga Rp117.520. Dapatkan di Shopee sekarang! https://t.co/rbJnZNORM3 https://t.co/hp6ZqlasQg	Check the Indonesian National Team Grade Jersey with a price of Rp117,520. Get it in Shopee now! https://t.co/rbjnzn0rm3 https://t.co/hp6zqiasqg
@KrnDi__ @aingriwehuy Kalo mau ngumpul lama suruh pada maen di liga Indonesia aja bang pemain timnas	@Krndi__ @aingriwehuy if you want to gather a long time to tell you to play in the Indonesian league, bang, the national team player
Striker naturalisasi milik Vietnam sempat masuk dalam daftar pemain termahal Asia Tenggara per Januari 2025. Namun pemain kelahiran Brasil tersebut kini justru sudah terlempar dari daftar 15 besar. https://t.co/ZGVNU0AWti	Vietnam's naturalized striker was included in the list of the most expensive Southeast Asia players as of January 2025. But the Brazilian -born player has now actually been thrown from the top 15 list. https://t.co/zgvnu0awti
Gawat! Timnas Indonesia Terancam	Bad! Indonesian national team threatened

#fypã,.viral #TRENDZ https://t.co/K2JK0iBh6J	
Polemik pelepasan pemain ke ASEAN All-Stars masih jadi topik perbincangan hangat khususnya di Vietnam. Kim Sang-sik benar-benar dibuat pusing PSSI pernyataan Erick Thohir dinilai merugikan ASEAN All-Stars. ðŸ†»ðŸ†³ https://t.co/24v70JWAKg	Polemic release of players to ASEAN All-Stars is still the topic of warm conversation, especially in Vietnam. Kim Sang-sik was really dizzy PSSI's statement Erick Thohir was considered detrimental to ASEAN All-Stars. ðŸ†»ðŸ†³ https://t.co/24v70JWAKG
Polemik pelepasan pemain ke ASEAN All-Stars masih jadi topik perbincangan hangat khususnya di Vietnam. Kim Sang-sik benar-benar dibuat pusing PSSI pernyataan Erick Thohir dinilai merugikan ASEAN All-Stars. ðŸ†»ðŸ†³ https://t.co/JCZQNqBcQN	Polemic release of players to ASEAN All-Stars is still the topic of warm conversation, especially in Vietnam. Kim Sang-sik was really dizzy PSSI's statement Erick Thohir was considered detrimental to ASEAN All-Stars. ðŸ†»ðŸ†³ https://t.co/jczqnqbcqn
@leleluyeah @worksfess Siapa tau bisa diapresiasi sebagai orang yang niat/pengen kerja kaya pelatih timnas Indonesia	@leleleleah @Worpsfess Who knows can be appreciated as someone who intention/wants to work rich in Indonesian national team coaches
PSSI melalui PT Garuda Sepak Bola Indonesia (GSI) menjalin kerja sama strategis dengan Djarum Foundation untuk menggelar turnamen sepak bola Piala Pertiwi U-14 dan U-16. Selengkapnya https://t.co/mR8s9WY7aN https://t.co/mR8s9WY7aN	PSSI through PT Garuda Football Indonesia (GSI) established a strategic cooperation with the Djarum Foundation to hold a U-14 and U-16 Pertiwi Cup soccer tournament. For complete https://t.co/mr8s9wy7an https://t.co/mr8s9wy7an
PSSI melalui PT Garuda Sepak Bola Indonesia (GSI) menjalin kerja sama strategis dengan Djarum Foundation untuk menggelar turnamen sepak bola Piala Pertiwi U-14 dan U-16. Selengkapnya https://t.co/5MQ3u6aEdi https://t.co/5MQ3u6aEdi	PSSI through PT Garuda Football Indonesia (GSI) established a strategic cooperation with the Djarum Foundation to hold a U-14 and U-16 Pertiwi Cup soccer tournament. For complete https://t.co/5mq3u6aedi https://t.co/5mq3u6aedi
@c4pedbaldy @rameankebulan @idextratime Mau soft atau ngga itu udh cukup menghentikan larinya Kvarats. Itu Klo timnas Indonesia yg digituin udh pasti penalti bang.	@c4pedbaldy @rameankebulan @idextratime want soft or not that it's enough to stop the running kvarats. That's if the Indonesian national team that is dawn that is definitely a penalty.
Ketua Umum PSSI Erick Thohir mengumumkan bahwa timnas Indonesia akan melaksanakan pemusatan latihan (TC) di Bali pada tanggal 26 Mei. ~MD #PSSI #ErickThohir #TimnasIndonesia https://t.co/1nFzMQ3sjb	PSSI Chairperson Erick Thohir announced that the Indonesian national team would carry out training camps (TC) in Bali on May 26. ~ Md #psi #erickthohir #timnasindonesia https://t.co/1NFZMQ3SJB
Gelandang Timnas Indonesia Joey Pelupessy merasakan perbedaan yang mencolok setelah kembali dari membela Skuad Garuda. ~MD #JoeyPelupessy	Indonesian national team midfielder Joey Pelupessy felt a striking difference after returning from defending the Garuda squad. ~ MD #Joeypelupessy

#TimnasIndonesia https://t.co/sMZ5i3KCx7	#timnasindonesia https://t.co/smz5i3kcx7
Ya tidak ada nama Hokky Caraka dalam daftar pemain PSS Sleman yang dibawa ke markas Persib Bandung pada pekan ke-30 Liga 1 2024-2025 kemarin. Dalam laga itu tim tamu kalah telak di GBLA. https://t.co/eSTQgrtHJl	Yes there is no name Hokky Caraka in the list of PSS Sleman players who were brought to the Persib Bandung headquarters in the 30th week of League 1 2024-2025 yesterday.In that match the visitors lost to a landslide at GBLA. https://t.co/estqgrthjl
Erick Thohir memastikan semua keputusan terkait pemanggilan pemain Timnas Indonesia ke skuad ASEAN All-Stars ada di tangan Patrick Kluivert. Seperti diketahui Jay Idzes dan kolega bakal menggelar TC di Bali pada 26 Mei mendatang. ðŸ†®ðŸ†© https://t.co/2v3uSZkMjH	Erick Thohir confirmed that all decisions related to the summons of the Indonesian national team players to the ASEAN All-Stars squad were in the hands of Patrick Kluivert.As is known Jay Idzes and colleagues will hold TC in Bali on May 26.ðŸ†®ðŸ†© https://t.co/2v3uszkmjh
Ya tidak ada nama Hokky Caraka dalam daftar pemain PSS Sleman yang dibawa ke markas Persib Bandung pada pekan ke-30 Liga 1 2024-2025 kemarin. Dalam laga itu tim tamu kalah telak di GBLA. https://t.co/UloWcgQqez	Yes there is no name Hokky Caraka in the list of PSS Sleman players who were brought to the Persib Bandung headquarters in the 30th week of League 1 2024-2025 yesterday.In that match the visitors lost to a landslide at GBLA. https://t.co/uiowcgqqez
Erick Thohir memastikan semua keputusan terkait pemanggilan pemain Timnas Indonesia ke skuad ASEAN All-Stars ada di tangan Patrick Kluivert. Seperti diketahui Jay Idzes dan kolega bakal menggelar TC di Bali pada 26 Mei mendatang. ðŸ†®ðŸ†© https://t.co/pwbh5hy0O1	Erick Thohir confirmed that all decisions related to the summons of the Indonesian national team players to the ASEAN All-Stars squad were in the hands of Patrick Kluivert.As is known Jay Idzes and colleagues will hold TC in Bali on May 26.ðŸ†®ðŸ†© https://t.co/pwbh5hy0o1
@prayoga2828 @garagarabola_ kan biasanya banyak oknum oknum yg berbicara bahwa banyaknya kiper asing di Indonesia atau kiper naturalisasi di Timnas jadi turun motivasinya wkwkwk biar ada alesan main jelek karena udh ga ada harepan dimainin di Timnas karen	@prayoga2828 @garagarabola_ There are usually many elements who speak that the number of foreign goalkeepers in Indonesia or naturalized goalkeepers in the national team so the motivation is down, so there is an ugly to play ugly because there is no harde
PEMAIN Bangkok United Pratama Arhan perlahan mulai kehilangan tempat di skuad utama Timnas Indonesia. Berita lengkapnya di https://t.co/aRGKPJRgN5 Sumber: https://t.co/L8HXOiLMYz https://t.co/VEcmKuMtt7	Bangkok United player Pratama Arhan slowly began to lose its place in the main squad of the Indonesian national team.Full news at https://t.co/argkplrjgg5 Source: https://t.co/l8hxoilmyz https://t.co/vecmkumtt7
Gawat! Timnas Indonesia terancam memainkan putaran 4 Kualifikasi Piala Dunia 2026 di Timur Tengah. Berita lengkapnya di https://t.co/e3GRZpq1ru Sumber: https://t.co/9lWg1gN64A https://t.co/P2XSGNLk0n	Bad!The Indonesian national team is threatened with playing round 4 of the 2026 World Cup qualifiers in the Middle East.Full news at https://t.co/e3grzpq1ru Source: https://t.co/9lw1gn64a https://t.co/p2xsgnlk0n

@utdfocusid @Bali_Football TC Timnas Indonesia Digelar Lebih Cepat Belajar dari Kekacauan saat Lawan Australia https://t.co/7oHpsgxpul	@utdfocusid @bali_football TC Indonesian National Team is held faster from chaos when opposed to Australia https://t.co/7oHPSGXPul
@aingriwehuy Gawat! Timnas Indonesia Terancam Mainkan Putaran 4 Kualifikasi Piala Dunia 2026 di Timur Tengah https://t.co/5MDDspmP4R	@aingriwehuy bad! Indonesian national team threatened to play round 4 of the 2026 World Cup qualifiers in the Middle East https://t.co/5mddspmp4r
@sorryyros @RaflIDP Arema telah memenangkan 2 gelar liga: 1992 (Galatama) dan 2009 10 (Indonesia Super League). Beberapa sumber menyebut 1 gelar karena debat soal gelar sebelum 1994. Arema FC adalah klub resmi saat ini dikelola PT Arema Aremania Bersatu Be	@Sorryyros @raflidp Arema has won 2 League titles: 1992 (Galatama) and 2009 10 (Indonesia Super League). Some sources mentioned 1 title because of a debate about the title before 1994. Arema FC was the official club currently managed by PT Arema Aremania Un
Timnas Indonesia Dirugikan Aturan AFC yang Gelar Babak Keempat Kualifikasi Piala Dunia 2026 Zona Asia di Timur Tengah? https://t.co/wWKppghNkL	The Indonesian national team was disadvantaged by the AFC rules, the fourth round of the 2026 Asian Zone in the Middle East? https://t.co/wwkppghnkl
Gawat! Timnas Indonesia terancam memainkan putaran 4 Kualifikasi Piala Dunia 2026 di Timur Tengah. Berita lengkapnya di https://t.co/e3GRZpq1ru Sumber: https://t.co/VLckM1hwYj https://t.co/bdnzyKNwnR	Bad! The Indonesian national team is threatened with playing round 4 of the 2026 World Cup qualifiers in the Middle East. Full news at https://t.co/e3grzpq1ru Source: https://t.co/vlckm1hwYj https://t.co/bdnzyknwnr
Penantian Kiper Cadangan di Bawah Sorot Lampu Stadion. Dalam sejarah Timnas Indonesia khususnya di era pelatih Patrick Kluivert para kiper yang menjadi pilihan utama maupun cadangan tentu memiliki tantangan tersendiri. https://t.co/N3zbSi9CYf https://t.co/mnsjQ1xnHA	Wait for a backup goalkeeper under the stadium lights. In the history of the Indonesian national team, especially in the era of coach Patrick Kluivert the goalkeepers who became the first choice and reserves certainly have their own challenges. https://t.co/n3zbsi9cyf https://t.co/mnsjq1xnha
Liga 1 2024/25 memasuki babak akhir yang akan menentukan nasib beberapa klub. PSSI berharap permintaan wasit asing jangan dilakukan mendadak. https://t.co/k1UIwPfeou	League 1 2024/25 enters the final round which will determine the fate of several clubs. PSSI hopes that foreign referees' demand will not be done suddenly. https://t.co/k1uiwpfeou
Ubah format di tengah jalan yalah bisa aje lu @afcasiancup So far cuma Timnas Indonesia aja nih yg bakal away jauh. Tim lainnya away ke negara tetangga. Kalo main di Iraq ngeri sama atmosfernya. Tapi kalo main di UAE ngeri juga. Owner Manchester City di negara ini cuy https://t.co/ngZr6C4sPm	Change the format in the middle of the road, you can, you can @afcasancup so far, only the Indonesian national team will be away. Other teams away to neighboring countries. If you play in Iraq horrified by the atmosphere. But if you play at UAE too horrified. Manchester City owner in this country cuy https://t.co/ngzr6c4spm
26 Pemain Timnas Putri Indonesia Siap Hadapi Yordania dan Bangladesh di FIFA Matchday Mei Mendatang -	26 Indonesian women's national team players are ready to face Jordan and Bangladesh at FIFA Matchday next May -

https://t.co/SYQdEZJxrY	https://t.co/syqdezjxry
Di musim 2025 jersey away Timnas Indonesia versi suporter hadir bak mahakarya sinematik paduan sempurna antara elegansi modern dan warisan budaya Nusantara. Didominasi warna putih bersih dengan aksen hitam dan abu-abu desain ini memancarkan aura kesederhanaan yang mewah. https://t.co/s7lJmxAtfJ	In the 2025 season Jersey Away Indonesian national team supporters were present like a cinematic mahakarya perfect alloy between modern elegance and the cultural heritage of the archipelago. Dominated by pure white with black and gray accents, this design emits a luxurious aura of simplicity. https://t.co/s7lJmxatfj
Ketua Umum PSSI Erick Thohir memastikan bahwa Timnas Indonesia akan menjalani pemusatan latihan (TC) lebih awal di Bali sebagai persiapan menghadapi dua laga krusial di ronde ketiga Kualifikasi Piala Dunia 2026 Grup C Lanjutan : https://t.co/23WiWqbb88 https://t.co/Y0NtSgk8py https://t.co/IsBmqiLLZ2	PSSI Chairperson Erick Thohir ensured that the Indonesian national team would undergo training camps (TC) earlier in Bali in preparation for two crucial matches in the third round of World Cup Qualifiers 2026 Group C Advanced: https://t.co/23wiwqbb88 https://t.co/y0ntsgk8py https://t.co/isbmqillz2
Ogah Ikut Campur soal Nasib Pratama Arhan di Timnas Indonesia Andre Rosiade Fokus ke Semen Padang #PratamaArhan #AndreRosiade #PatrickKluivert #TimnasIndonesia #SkuadGaruda #TiMerahPutih #SemenPadang #KualifikasiPialaDunia2026 https://t.co/Sd2nJUiDQb lewat @tribunnews	Unwilling to interfere with the fate of Pratama Arhan in the Indonesian National Team Andre Rosiade focused on Semen Padang #PratamaArhan #Andrerosiade #Patrickkluivert #TimnasIndonesia #squadgaruda #Timerahputih #semenpadang #qualificationpialadunia2026 https://t.co/sd2njuidqb through @tribunnews
>> RUMOR << Cahya Supriyadi di incar oleh Laskar Mataram. Kiper 22 tahun ini memang sudah saatnya kembali ke Liga 1. Musim lalu Cahya tampil bagus bersama FC Bekasi City sehingga tetap dipanggil Timnas Indonesia meski bermain di Liga 2. https://t.co/J8gxrnG4J1	>> Rumors << Cahya Supriyadi was targeted by Laskar Mataram. This 22 -year -old goalkeeper is indeed time to return to League 1. Last season Cahya performed well with FC Bekasi City so that it was still called the Indonesian national team despite playing in League 2. https://t.co/j8gxrng4j1
Marc Klok sukses membuktikan kapasitasnya sebagai salah satu pemain penting Persib Bandung. Gelandang berlabel Timnas Indonesia itu mengakui tidak diam saja setelah menerima kritikan Bobotoh. https://t.co/sa3fpToOHQ	Marc Klok successfully proves his power as one of the important players of Persib Bandung. The midfielder labeled the Indonesian national team admitted that he was not silent after receiving Bobotoh's criticism. https://t.co/sa3fptooHQ
Marc Klok sukses membuktikan kapasitasnya sebagai salah satu pemain penting Persib Bandung. Gelandang berlabel Timnas Indonesia itu mengakui tidak diam saja setelah menerima kritikan Bobotoh. https://t.co/bNsNPFitou	Marc Klok successfully proves his power as one of the important players of Persib Bandung. The midfielder labeled the Indonesian national team admitted that he was not silent after receiving Bobotoh's criticism. https://t.co/bnsnpfitou

Erick Thohir Ungkap Penyebab Belum Tunjuk Pelatih Timnas Indonesia untuk SEA Games 2025 https://t.co/uasrCz3pQn	Erick Thohir revealed the cause of the Indonesian National Team Coach for the 2025 SEA Games https://t.co/uasrcz3pqn
OPEN LIST TIKET TIMNAS KLOTER 1 SAMPE 2 MEI 2025 INDONESIA VS CHINA 05 JUNI 2025 READY SEMUA KATEGORI Bisa cod lok jaktim fast respon #zonauang #zonajajani, 2i, 2 wts wtb	Open list of national team ticket tickets 1 to May 2, 2025 Indonesia vs China 05 June 2025 Ready All categories can be cod jaktim fast response #zonauang #zonajajani, 2i, 2 WTS WTB
Ole RomenyÂ menilai bahwa keputusan bergabung keÂ timnas IndonesiaÂ sudah tepat. OleÂ menjelaskan bahwa bergabung keÂ timnas IndonesiaÂ memiliki kebanggaan tersendiri https://t.co/xptRKdj4oz	Ole Romeny considered that the decision to join the Indonesian national team was right. Ole explained that joining the Indonesian national team had its own prizes https://t.co/xptrkdj4oz
Ole RomenyÂ menilai bahwa keputusan bergabung keÂ timnas IndonesiaÂ sudah tepat. OleÂ menjelaskan bahwa bergabung keÂ timnas IndonesiaÂ memiliki kebanggaan tersendiri https://t.co/6HESJP434f	Ole Romeny considered that the decision to join the Indonesian national team was right. Ole explained that joining the Indonesian national team had its own prizes https://t.co/6HesJP434F
Akhir pekan lalu bek Timnas Indonesia Rizky Ridho unjuk kebolehan di lapangan bola dalam sebuah ajang mini soccer. Sekadar friendly sih. Tapi tetap seru! https://t.co/XePHUsqJU9	Last weekend Indonesian national team defender Rizky Ridho showed his skill on the soccer field in a mini soccer event. Just friendly anyway. But still exciting! https://t.co/xephusqju9
Pelatih Malaysia mengungkapkan cara mendongkrak prestasi timnya yang tidak akan mencontoh PSSI. Seperti diketahui Timnas Indonesia berkembang pesat berkat program naturalisasi yang berjalan dengan baik. ðŸ†®ðŸ†©ðŸ†²ðŸ†¼ https://t.co/dGqKZyThG0	Malaysian coach revealed how to boost the achievements of his team that would not imitate PSSI. As is known the Indonesian national team is developing rapidly thanks to the well-run naturalized program. ðŸ†®ðŸ†©ðŸ†²ðŸ†¼ https://t.co/dgqkzythg0
Pelatih Malaysia mengungkapkan cara mendongkrak prestasi timnya yang tidak akan mencontoh PSSI. Seperti diketahui Timnas Indonesia berkembang pesat berkat program naturalisasi yang berjalan dengan baik. ðŸ†®ðŸ†©ðŸ†²ðŸ†¼ https://t.co/ETEDL68xAW	Malaysian coach revealed how to boost the achievements of his team that would not imitate PSSI. As is known the Indonesian national team is developing rapidly thanks to the well-run naturalized program. ðŸ†®ðŸ†©ðŸ†²ðŸ†¼ https://t.co/etedl68xaw
Cek Jersey Timnas Indonesia dengan harga Rp119.328. Dapatkan di Shopee sekarang! https://t.co/3UHYkBOcjF https://t.co/CUDRBPngrW	Check the Indonesian national team jersey at a price of Rp119,328. Get it in Shopee now! https://t.co/3uhykbocjf https://t.co/cudrbpngrw
Berdasarkan keterangan terkini dari Ketum PSSI Erick Thohir kepada awak media tak hanya Ferarri Asnawi pun masih diragukan untuk ikut serta dalam skuad ASEAN All	Based on the latest information from PSSI Chairman Erick Thohir to the media crew not only Ferrarri Asnawi is still doubtful to participate in the ASEAN All Stars squad to

Stars guna melakoni laga persahabatan melawan Man Utd https://t.co/atdy0keb0U	play a friendly match against Man Utd https://t.co/atdy0keb0U
Berdasarkan keterangan terkini dari Ketum PSSI Erick Thohir kepada awak media tak hanya Ferarri Asnawi pun masih diragukan untuk ikut serta dalam skuad ASEAN All Stars guna melakoni laga persahabatan melawan Man Utd https://t.co/w3bi5TmHbv	Based on the latest information from PSSI Chairman Erick Thohir to the media crew not only Ferrarri Asnawi is still doubtful to participate in the ASEAN All Stars squad to play a friendly match against Man Utd https://t.co/w3bi5TmHbv
AFF dan Sea Games di luar agenda FIFA Skuad Timnas U-23 cukup pakai pemain Liga 1 saja Jgn panggil utk ganggu pemain naturalisasi berlatih di Klub nya masing2 PSSI Erick Thohir #KitaGaruda @erickthohir @PSSI Timnas Indonesia https://t.co/0DzVfQbpHD	AFF and SEA Games outside the FIFA U-23 national team squad agenda just use League 1 players, don't call to disturb naturalization players to train at their clubs Erick Thohir #kitagaruda @erickthohir @psi Indonesian National Team https://t.co/0dzvfqbphd
Daftar 26 Pemain Timnas Putri Indonesia vs Yordania dan Bangladesh https://t.co/ZF4frzZBNi	List of 26 Indonesian women's national team players vs Jordan and Bangladesh https://t.co/zf4frzbni
Ali Jasim Mimpi Buruk Timnas Indonesia yang Kini Jadi Simbol Persaudaraan Bersama Thom Haye! https://t.co/qgfu4nghAG	Ali Jasim Nightmare of the Indonesian National Team who is now a symbol of brotherhood with Thom Haye! https://t.co/qgfu4nghag
Patrick Kluivert pelatih Timnas Indonesia telah membuat langkah mengejutkan dengan memanggil empat pemain naturalisasi tambahan untuk laga melawan China dan Jepang. https://t.co/wwNFsl4cai	Patrick Kluivert Indonesian national team coach has made a surprising step by calling four additional naturalized players for the match against China and Japan. https://t.co/wwnfs14cai
Patrick Kluivert pelatih Timnas Indonesia telah membuat langkah mengejutkan dengan memanggil empat pemain naturalisasi tambahan untuk laga melawan China dan Jepang. https://t.co/GpWGFbaOJ2	Patrick Kluivert Indonesian national team coach has made a surprising step by calling four additional naturalized players for the match against China and Japan. https://t.co/gpwgfbaoj2
Patrick Kluivert pelatih Timnas Indonesia telah membuat langkah mengejutkan dengan memanggil empat pemain naturalisasi tambahan untuk laga melawan China dan Jepang. https://t.co/YMXi85AP7o	Patrick Kluivert Indonesian national team coach has made a surprising step by calling four additional naturalized players for the match against China and Japan. https://t.co/ymxi85ap7o
#kesenjangan sosial Cek Jersey Bola Timnas Indonesia dengan harga Rp117.520. Dapatkan di Shopee sekarang! https://t.co/qPR4EVhFNw https://t.co/Cborf5uNMs	#Social Association Check Indonesian National Team Football Jersey at a price of Rp117,520. Get it in Shopee now! https://t.co/qpr4evhfnw https://t.co/cborf5unms
Klasemen Sementara Calon Tuan Rumah Babak Keempat Kualifikasi Piala Dunia 2026 Zona Asia: Timnas Indonesia Urutan	Standings for the fourth round of the 2026 Asian Zone Qualifiers: Indonesian National Team Bloated? https://t.co/n7u4yps3k6

Buncit? https://t.co/N7u4yps3k6	
@japra97 @unilubis Ternyata pemain keturunan selama ini dibayar supaya mau gabung timnas Indonesia pakai duit siapa hayo? Ya pakai duit BUMN lahbayar agen dan bayar pemain gak mungkin sekelas Kevin Diks Martin Paes Mees Hilgers mau bela timnas kalau cuma	@japra97 @unilubis it turns out that descendants have been paid so that they want to join the Indonesian national team using money, hi?Yes, use the BUMN Lahbayar agent and pay the players, it is impossible for the class of Kevin Diks Martin Paes Mees Hilg
@BudiBukanIntel ekonomi menengah kebawah yang terlibat di ekosistem sepakbola usia muda di Indonesia. Beruntung PSSI di kepengurusan saat ini masih memberi kesempatan besar terhadap pemain muda SSB/Akademi yang dikelola masyarakat ekonomi menengah kebawah	@Budempitintel Middle and Lower Economy involved in a young soccer ecosystem in Indonesia.Luckily PSSI in the management currently still provides a great opportunity to young SSB players/academies managed by the middle to lower economic community
Timnas Indonesia Bakal Lakukan TC di Bali Erick Thohir Sebut Belum Pasti Lepas Pemain ke ASEAN All Star https://t.co/7X2XOFRvDx	Indonesian national team will do TC in Bali Erick Thohir Mentioned not yet released players to ASEAN All Star https://t.co/7x2xofrvdx
@BudiBukanIntel berjuang secara mandiri dan bertahap apalagi mereka lebih kasian melihat pelatih asing yg melatih Timnas K.U yang harus lebih berusaha keras untuk meningkatkan potensi pesepakbola Indonesia (Seperti BA Produk Minuman) dibanding berempati k	@Budentiintel struggles independently and gradually especially if they are more sorry to see foreign coaches who train the K.U national team who must be more trying hard to increase the potential of Indonesian footballers (such as BA Beverage products) ra
@BudiBukanIntel UNTUK NRIMO ketika anak didiknya yang berhasil terpantau masuk Timnas Indonesia K.U (Lewat Soeratin atau EPA) statusnya dianggap pemain sementara (dibanding memaksimalkan potensinya di TC TIMNAS) dan harus terima untuk posisinya di Timnas	@Budentipintel for Nrimo when his students who were successfully monitored to enter the Indonesian national team K.U (through Soeratin or EPA) their status is considered a temporary player (compared to maximizing his potential in the national team TC) and
@BudiBukanIntel Masyarakat kelas ekonomi menengah kebawah yang terlibat di ekosistem sepakbola usia muda di Indonesia yang berjuang secara mandiri dan bertahap DITUNTUT UNTUK tidak berharap pemain didikan nya bisa dipantau langsung masuk Timnas Indonesia	@Budembinintel Middle Economy Class Community Lower Involved in a Young Age Football Ecosystem in Indonesia who fought independent
Garuda memanggil! Berikut adalah skuad Timnas Indonesia untuk FIFA Matchday akhir Mei ini kontra Yordania dan Bangladesh di Yordania. ĩ, 31 Mei s/d 3 Juni https://t.co/HWVppjHwyg	Garuda calls!The following is the Indonesian national team squad for FIFA Matchday at the end of this May against Jordan and Bangladesh in Jordan.ĩ, 31 to June 3 https://t.co/hwvppjhwyg
TC Timnas Indonesia Bentrok dengan	Indonesian national team tc clashes with

ASEAN All Stars vs Manchester United Erick Thohir Bicara Nasib 2 Pemain Ini https://t.co/NeSQGK8fwP	ASEAN All Stars vs Manchester United Erick Thohir Talking
Jakarta Garuda Jaya yang merupakan Timnas Voli Putra Indonesia U-21 melakukan uji coba dengan mengikuti turnamen Win Streak Volleyball Invitational di Chinese Taipei Semoga mereka bisa maksimal kesempatan ini menuju Kejuaraan FIVB U-21 pada bulan Agustus nanti https://t.co/L2NU9fl9vA	Jakarta Garuda Jaya, which is the Indonesian men's volleyball national team U-21 conducted a trial by taking part in the Win Streak Volleyball Invitational tournament at Chinese Taipei, hopefully they can maximize this opportunity to the FIVB U-21 championship in August https://t.co/l2nu9fl9va
Gawat! Timnas Indonesia Terancam Mainkan Putaran 4 Kualifikasi Piala Dunia 2026 di Timur Tengah https://t.co/5MDDspmhfj	Bad! Indonesian national team threatened to play round 4 of the 2026 World Cup qualifiers in the Middle East https://t.co/5mddspmhfj
Mana nih Sport Lovers yang kemaren nanyain tentang hal ini Nah ini dia alasan kenapa Coach Indra Sjafri masih kerja di PSSI walau udah dipecat sebagai pelatih Timnas Indonesia #TimnasIndonesia #IndraSjafri #PSSI https://t.co/rQUwD2J8il	Where is the Sport Lovers that yesterday asked about this, this is the reason why Coach Indra Sjafri is still working at PSSI even though it has been fired as the Indonesian national team coach
Ketua Umum PSSI Erick Thohir memastikan pemusatan latihan atau TC Timnas Indonesia akan dimulai pada 26 Mei 2025 di Bali untuk Putaran Ketiga Kualifikasi Piala Dunia 2026 Zona Asia. https://t.co/MJ5bVrVJH1	PSSI Chairperson Erick Thohir ensured that the training camp or TC Indonesian national team will begin on May 26, 2025 in Bali for the third round of the 2026 Asian Zone Qualifiers. https://t.co/mj5bvrvjh1
Performa Marselino Ferdinan Bersama Oxford United Meningkatkan Apakah Bintang Timnas Indonesia Ini Akan Bertahan di Liga Inggris? #marselinoferdinan #oxfordunited #ligainggris https://t.co/UHSgTP43me	Marselino Ferdinan's performance with Oxford United increases whether the Indonesian national team star will survive in the English League? #marsselinailferdinan #oxfordunited #ligainggris https://t.co/uhsghp43me
China Ajukan Jadi Tuan Rumah Babak Keempat Kualifikasi Piala Dunia 2026 Zona Asia Yakin Singkirkan Timnas Indonesia? https://t.co/43tSEQbMw5	China proposed to host the fourth round of the 2026 World Cup Qualifiers Asia Zone sure to get rid of the Indonesian national team? https://t.co/43tseqbmw5
Kejutan... ini Sederet Nama Pemain Naturalisasi Terbaru yang Bakal Perkuat Timnas Indonesia Saat Lawan China? https://t.co/wSe9YH7WGT	Surprise ... This is a series of names of the latest naturalized players who will strengthen the Indonesian national team when opponents China? https://t.co/wse9yh7wgt
Joey Pelupessy merasakan atmosfer luar biasa dari suporter yang rela datang mendukung langsung Tim Merah Putih. Pemain Lommel SK itu bahkan mengaku tak pernah membayangkan stadion bakal dipenuhi suporter	Joey Pelupessy felt the extraordinary atmosphere of supporters who were willing to come directly to support the Red and White team. The Lommel SK player even claimed he never imagined the stadium would be filled with supporters

https://t.co/sQuEXxkOJ6	https://t.co/squexxkoj6
Joey Pelupessy merasakan atmosfer luar biasa dari suporter yang rela datang mendukung langsung Tim Merah Putih. Pemain Lommel SK itu bahkan mengaku tak pernah membayangkan stadion bakal dipenuhi suporter https://t.co/cpj0N7VCw1	Joey Pelupessy felt the extraordinary atmosphere of supporters who were willing to come directly to support the Red and White team. The Lommel SK player even claimed he never imagined the stadium would be filled with supporters https://t.co/cpj0N7vcw1
BOLAVITA NEWS Gelandang Timnas Indonesia Thom Haye mengungkapkan bahwa timnya akan menggelar pemusatan latihan di Bali untuk melawan Timnas China dan Timnas Jepang. Dukung akun ini melalui link yang ada di bio ya https://t.co/vkqxW9hRau	Bolavita News Indonesian national team midfielder Thom Haye revealed that his team will hold training camps in Bali to fight the Chinese national team and the Japanese national team. Support this account through the link in Bio https://t.co/vkqxxw9hrau
Saat program tim nasional ini mulai dijalankan dua tahun lalu Timnas Wanita Indonesia hanya memiliki tujuh pemain senior. Kini jumlah tersebut sudah bertambah menjadi 15 orang dan ditargetkan bisa mencapai 20 pemain dalam waktu dekat Hmmm...	When the national team program began two years ago the Indonesian women's national team only had seven senior players. Now the number has increased to 15 people and is targeted to reach 20 players in the near future Hmmm ...
Joey Pelupessy Ungkap Perbedaan Kontras yang Dirasakan Usai Bela Timnas Indonesia https://t.co/WBF0WOHZGZ	Joey Pelupessy revealed the difference in contrast felt after defending the Indonesian national team https://t.co/wbf0wohzgz
Semula ada tiga pemain berlabel timnas Indonesia yang bergabung bersama Status Pro. Mereka adalah Marc Klok Sandy Walsh dan Kadek Arel https://t.co/reo09JqC2C	Initially there were three players labeled the Indonesian national team who joined the Pro status. They are Marc Klok Sandy Walsh and Kadek Arel https://t.co/reo09jqc2c
Semula ada tiga pemain berlabel timnas Indonesia yang bergabung bersama Status Pro. Mereka adalah Marc Klok Sandy Walsh dan Kadek Arel https://t.co/gLyN9Z9P3y	Initially there were three players labeled the Indonesian national team who joined the Pro status. They are Marc Klok Sandy Walsh and Kadek Arel https://t.co/glyn9z9p3y
Dibocorkan Sang Istri Soal Naturalisasi Ciro Alves Siap Perkuat Timnas Indonesia? https://t.co/uwNvYyzgC2	Leaked by his wife about the naturalization of Ciro Alves ready to strengthen the Indonesian national team? https://t.co/uwnvyyzgc2

Lampiran 3. Hasil Pelabelan Dengan VADER

clean_text	Sentiment
indonesian national team plans face malaysia lebanon september	Neutral
ole romeny want defend indonesian national team thanks garuda fans	Positive
indonesian national team assistant coach denny landzaat targeted willem ii coach	Neutral
indonesian womens national team announced names fifa matchday jordanian bangladeshso preparation asean womens championship	Neutral
pratama arhan began lose place indonesian national team azizah salshas father said	Negative
additional naturalization players championed patrick kluivert indonesian national team match vs china japan	Positive
like indonesian national team defense years ago meet arab team	Positive
details best offer thursday may indonesian national team home size l x cm brand new tag nns remake micro hole barely visible left back smoothly handsome ieu bin k best offer	Positive
alex pastoor removing agenda indonesian national team vs russia new polemic denis manturov spoke	Neutral
detail best offer thursday may indonesian national team official parka jacket size x cm brand new smooth tags none minusnote commercial sold produced players uselimited rarebin k best offer	Positive
used indonesian national team assistant	Neutral
indonesian national team hosting fourth round asian world cup qualifiers meet condition	Neutral
patrick kluivert calls ciro alves indonesian national team officially become citizen	Neutral
saving feel comfort style indonesian national team soccer jersey	positive
interested knowing salaries indonesian u national team playersone mainstay players iqbal gwijangge known receive fairly fantastic salarycheck full article	positive
indonesian national team coach son shane kluivert trophy uyl	neutral
looking forngebon players strangle opportunity local players	positive
pratama arhan retainer garuda indonesian national team squad finished match club bangkok united continued league thailand week amp	positive
check indonesian national team football jersey dryfit adult price rpget shopee	neutral
check indonesian national team jersey ori dryfit adult grade price rpget shopee	neutral
erick thohir rejected idea becoming general chair pssi life	negative
prepare welcome brighter future freeport grassroots tournament becomes forum help formation character ability indonesian childrens football early agewatch full show channel pssi	positive
ok means xabi trains indonesian national team wkwk	positive

standings fourth round asian zone qualifiers indonesian national team bloated	neutral
pssi determine indonesian u national team coach sea games exco meeting	neutral
soccer national team actually follow done massive basketball national team strengthens national team outside players especially indonesian blood	positive
indonesian national team rumored holding international tournament septemberlater countries invited	neutral
countries appear th round asian world cup qualifiers compete automatic qualify ticketsthe united arab emirates iraq favored hostwhat indonesian national team	positive
indonesian putri u national team joined group asian cup qualificationthe inaugural match india held august myanmar	neutral
indonesian national team rumored holding international tournament septemberlater countries invited	neutral
indonesia nothing possible bro money conditioned remember fake doctor became medical team pssi national team less absurd try	negative
countries appear th round asian world cup qualifiers compete automatic qualify ticketsthe united arab emirates iraq favored hostwhat indonesian national team	positive
indonesian national team assistant coach denny landzaat targeted willem ii coach	neutral
pratama arhan began lose place indonesian national team azizah salshas father said	negative
robi darwiss salary defending midfielder indonesian national team curiousstake peek much income shining young player	neutral
indonesian national team defensive midfielder rivaldo pakpahan spotlightbut much salary nowcheck details article	positive
joey pelupessey called features indonesian national team squad patrick kluivert	neutral
total pss sleman players absent persib bandung th week league player returned training pieter huistra foster children	neutral
total pss sleman players absent persib bandung th week league player returned training pieter huistra foster children	neutral
ole romeny recalled goal indonesian national team striker indonesian national team ole romeny recalled special moment garuda team made debut appearance goal debut facing australia sydney stadium marchquoted official page	positive
latest information rumors jay idzes transfer april based previous data jay idzes venezia players indonesian national teams associated clubs inter milan bolognahis appearance se	neutral
unfortunately indonesian national team next fifa matchday still without Shalika	negative
afc changed format round world cup qualifiersindonesian national team benefit	positive
market value indonesian national team players highest rp billion	positive
gpp middle east also many indonesians live still militant support national team	positive

check indonesian national team grade jersey price rpget shopee	neutral
want gather long time tell play indonesian league bang national team player	positive
vietnams naturalized striker included list expensive southeast asia players january brazilian born player actually thrown top list	positive
badindonesian national team threatened play round world cup qualifiers middle east	negative
vietnams naturalized striker included list expensive southeast asia players january brazilian born player actually thrown top list	positive
joey pelulessys confession debut indonesian national team know taste different	neutral
astaga ole romeny indonesian football player	neutral
provides complete games bonuses every day	positive
wonderkid china said secret weapon indonesian national team	negative
persija goalkeeper data agus waluyo indonesian national team	neutral
kan fmd september october novemberstill want workwell must proposed afcdont surrender told play one placewhere play middle east indonesia least	positive
antv tobali viral	neutral
antv tobali viral	neutral
polemic release players asean allstars still topic warm conversation especially vietnamkim sangsik really dizzy pssis statement erick thohir considered detrimental asean allstars	negative
polemic release players asean allstars still topic warm conversation especially vietnamkim sangsik really dizzy pssis statement erick thohir considered detrimental asean allstars	negative
knows appreciated someone intentionwants work rich indonesian national team coaches	positive
pssi pt garuda football indonesia gsi established strategic cooperation djarum foundation hold u u pertiwi cup soccer tournamentfor complete	neutral
pssi pt garuda football indonesia gsi established strategic cooperation djarum foundation hold u u pertiwi cup soccer tournamentfor complete	neutral
want soft enough stop running kvaratsthats indonesian national team dawn definitely penalty	negative
pssi chairperson erick thohir announced indonesian national team would carry training camps tc bali may md	neutral
indonesian national team midfielder joey pelulessy felt striking difference returning defending garuda squad md	neutral
yes name hokky caraka list pss sleman players brought persib bandung headquarters th week league yesterdayin match visitors lost landslide gbla	negative
erick thohir confirmed decisions related summons indonesian national team players asean allstars squad hands patrick kluivertas known jay idzes colleagues hold tc bali may	neutral
yes name hokky caraka list pss sleman players brought persib bandung headquarters th week league yesterdayin match visitors lost landslide gbla	negative
erick thohir confirmed decisions related summons indonesian national team players asean allstars squad hands patrick kluivertas known jay idzes colleagues hold tc bali may	neutral

usually many elements speak number foreign goalkeepers indonesia naturalized goalkeepers national team motivation ugly play ugly harde	negative
bangkok united player pratama arhan slowly began lose place main squad indonesian national teamfull news source	neutral
badthe indonesian national team threatened playing round world cup qualifiers middle eastfull news source	negative
tc indonesian national team held faster chaos opposed australia	negative
badindonesian national team threatened play round world cup qualifiers middle east	negative
arema league titles galatama indonesia super leaguesome sources mentioned title debate title arema fc official club currently managed pt arema aremania united	positive
indonesian national team disadvantaged afc rules fourth round asian zone middle east	negative
badthe indonesian national team threatened playing round world cup qualifiers middle eastfull news source	negative
wait backup goalkeeper stadium lightsin history indonesian national team especially era coach patrick kluivert goalkeepers became first choice reserves certainly challenges	positive
league enters final round determine fate several clubspssi hopes foreign referees demand done suddenly	positive
change format middle road far indonesian national team awayother teams away neighboring countriesif play iraq horrified atmospherebut play uae horrifiedmanchester city owner country cuy	positive
indonesian womens national team players ready face jordan bangladesh fifa matchday next may	positive
season jersey away indonesian national team supporters present like cinematic mahakarya perfect alloy modern elegance cultural heritage archipelagodominated pure white black gray accents design emits luxurious aura simplicity	positive
pssi chairperson erick thohir ensured indonesian national team would undergo training camps tc earlier bali preparation two crucial matches third round world cup qualifiers group c advanced wiwqbb	positive
unwilling interfere fate pratama arhan indonesian national team andre rosiade focused semen padang	positive
gt gtrumors lt itcahya supriyadi targeted laskar mataramthis year old goalkeeper indeed time return league last season cahya performed well fc bekasi city still called indonesian national team despite playing league	positive
marc klok successfully proves power one important players persib bandungthe midfielder labeled indonesian national team admitted silent receiving bobotohs criticism	positive
marc klok successfully proves power one important players persib bandungthe midfielder labeled indonesian national team admitted silent receiving bobotohs criticism	positive
erick thohir revealed cause indonesian national team coach sea games	neutral
open list national team ticket tickets may indonesia vs china june ready categories cod jaktim fast response wts wtbb	positive
ole romeny considered decision join indonesian national team rightole	positive

explained joining indonesian national team prizes	
ole romeny considered decision join indonesian national team rightole explained joining indonesian national team prizes	positive
last weekend indonesian national team defender rizky ridho showed skill soccer field mini soccer eventjust friendly anywaybut still exciting	positive
malaysian coach revealed boost achievements team would imitate pssias known indonesian national team developing rapidly thanks well run naturalized program	positive
malaysian coach revealed boost achievements team would imitate pssias known indonesian national team developing rapidly thanks well run naturalized program	positive
check indonesian national team jersey price rpget shopee	neutral
based latest information pssi chairman erick thohir media crew ferrarri asnawi still doubtful participate asean stars squad play friendly match man utd	positive
based latest information pssi chairman erick thohir media crew ferrarri asnawi still doubtful participate asean stars squad play friendly match man utd	positive
aff sea games outside fifa u national team squad agenda use league players dont call disturb naturalization players train clubs erick thohir indonesian national team	positive
list indonesian womens national team players vs jordan bangladesh	neutral
ali jasim nightmare indonesian national team symbol brotherhood thom haye	neutral
patrick kluivert indonesian national team coach made surprising step calling four additional naturalized players match china japan	positive
patrick kluivert indonesian national team coach made surprising step calling four additional naturalized players match china japan	positive
patrick kluivert indonesian national team coach made surprising step calling four additional naturalized players match china japan	positive
association check indonesian national team football jersey price rpget shopee	neutral
standings fourth round asian zone qualifiers indonesian national team bloated	neutral
turns descendants paid want join indonesian national team using money hiyes use bumh lahbayar agent pay players impossible class kevin diks martin paes mees hilgers defend national team	positive
middle lower economy involved young soccer ecosystem indonesialuckily pssi management currently still provides great opportunity young ssb playersacademies managed middle lower economic community	positive
indonesian national team tc bali erick thohir mentioned yet released players asean star	neutral
struggles independently gradually especially sorry see foreign coaches train ku national team must trying hard increase potential indonesian footballers ba beverage products rather approaching k	negative
nrimo students successfully monitored enter indonesian national team ku soeratin epa status considered temporary player compared maximizing potential national team tc must accept position national team	positive

middle economy class community lower involved young age football ecosystem indonesia fought independent	negative
garuda callsthe following indonesian national team squad fifa matchday end may jordan bangladesh jordan may june	neutral
indonesian national team tc clashes asean stars vs manchester united erick thohir talking	Positive
jakarta garuda jaya indonesian mens volleyball national team u conducted trial taking part win streak volleyball invitational tournament chinese taipei hopefully maximize opportunity fivb u championship august	Positive
badindonesian national team threatened play round world cup qualifiers middle east	Negative
sport lovers yesterday asked reason coach indra sjafri still working pssi even though fired indonesian national team coach	Negative
pssi chairperson erick thohir ensured training camp tc indonesian national team begin may bali third round asian zone qualifiers	Neutral
marselino ferdinans performance oxford united increases whether indonesian national team star survive english league	Positive
china proposed host fourth round world cup qualifiers asia zone sure get rid indonesian national team	Positive
surprise series names latest naturalized players strengthen indonesian national team opponents china	Positive
joey pelulessy felt extraordinary atmosphere supporters willing come directly support red white teamthe lommel sk player even claimed never imagined stadium would filled supporters	Positive
joey pelulessy felt extraordinary atmosphere supporters willing come directly support red white teamthe lommel sk player even claimed never imagined stadium would filled supporters	positive
bolavita news indonesian national team midfielder thom haye revealed team hold training camps bali fight chinese national team japanese national teamsupport account link bio	negative
national team program began two years ago indonesian womens national team seven senior playersnow number increased people targeted reach players near future hmmm	positive
joey pelulessy revealed difference contrast felt defending indonesian national team	neutral
initially three players labeled indonesian national team joined pro statusthey marc klok sandy walsh kadec arel	neutral
initially three players labeled indonesian national team joined pro statusthey marc klok sandy walsh kadec arel	neutral