

## Analisis Sentimen Masyarakat Pengguna Media Sosial Twitter Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT)

Nelly Sofi<sup>1</sup>, Tri Sulistyorini<sup>2</sup>, Muhammad Nazaruddin<sup>3</sup>

<sup>1,2,3</sup> Program Studi Informatika, Universitas Gunadarma

### Abstrak

Balapan MotoGP Satu di Nusa Tenggara Barat Lombok, Mandalika yang diselenggarakan pada 18 Maret 2022, mendapatkan banyak tanggapan ataupun teaksi dari masyarakat di media social terutama *Twitter*. Tanggapan-tanggapan tersebut ada yang setuju dan tidak mengenai penyelenggaraan MotoGP di Mandalika ini, untuk mengetahui tanggapan masyarakat yang setuju atau tidak diperlukan sistem yang dapat mengolah data *tweets* dengan metode analisis sentimen. Penggunaan BERT (*Bidirectional Encoder Representations from Transformers*) untuk analisis sentimen menghasilkan sebuah model bahasa dua arah (*bidirectional*) yang dapat memahami konteks keseluruhan kata dari sebuah kalimat. *Dataset* yang digunakan melalui tahap preprocessing seperti *case folding*, *data cleaning*, tokenisasi, normalisasi, dan penghapusan *stopwords* sebelum dilakukan analisis sentimen. Penelitian ini menggunakan beberapa *hyperparameters* yaitu *batch size* sebesar 32, *optimizer* menggunakan *Adam* dengan *learning rate*  $3e-6$  atau 0.000003, dan *epoch* 25. Hasil evaluasi dari model tersebut mendapatkan akurasi sebesar 55%. *Precision* untuk positif sebesar 56%, netral sebesar 59%, dan negatif sebesar 44%. *Recall* untuk positif sebesar 74%, netral sebesar 29%, dan negatif 54%. *F1-score* untuk positif sebesar 64%, netral sebesar 38%, dan negatif sebesar 48%.

**Kata kunci:** Analisis Sentimen, MotoGP, Mandalika, Lombok, BERT.

### Abstract

*The MotoGP One race in West Nusa Tenggara Lombok, Mandalika which was held on March 18 2022, received many responses or reactions from the public on social media, especially Twitter. There are those who agree and disagree about the holding of MotoGP in Mandalika, to find out the responses of the people who agree or disagree is needed that can process tweets data using the sentiment analysis method. The use of BERT (Bidirectional Encoder Representations from Transformers) for sentiment analysis produces a bidirectional language model that can understand the context of all words from a sentence. The dataset used goes through preprocessing stages such as case folding, data cleaning, tokenization, normalization, and removal of stopwords before sentiment analysis is carried out. This study uses several hyperparameters, namely a batch size of 32, the optimizer uses Adam with a learning rate of  $3e-6$  or 0.000003, and an epoch of 25. The evaluation results of the model obtain an accuracy of 55%. Precision for positive by 56%, neutral by 59%, and negative by 44%. Recall for positive is 74%, neutral is 29%, and negative is 54%. F1-score for positive is 64%, neutral is 38%, and negative is 48%.*

**Keywords:** Sentiment Analysis, MotoGP, Mandalika, Lombok, BERT.

## **PENDAHULUAN**

Presiden Indonesia Jokowi Dodo meresmikan pembukaan Pertamina Mandalika International Street Circuit pada tanggal 12 November 2021, yang beralokasikan pada Provinsi Nusa Tenggara Barat (NTB) daerah Lombok. Pembangunan Sirkuit Mandalika sendiri sudah dimulai sejak akhir tahun 2019, yang total biaya pembangunan Sirkuit Mandalika mencapai Rp 1,1 triliun. Dengan diresmikannya sirkuit berstandar internasional tersebut telah menyelenggarakan beberapa acara berskala Internasional seperti ajang balap motor MotoGP, yang berhasil membawa banyak manfaat dalam segi perekonomian dan berpotensi menghasilkan 4,5 triliun rupiah.

Indonesia menjadi tuan rumah dari dua event olahraga balap internasional yaitu MotoGP, yang mana merupakan olahraga balap motor bergengsi dan dikenali banyak kalangan masyarakat. Indonesia menggelar ajang MotoGP 2022 pada tanggal 18-20 Maret 2022, dimana para rider kelas dunia seperti Marc Marquez, Fabio Quartararo, Miguel Oliveira, dan lainnya akan bersaing merebut podium juara. Event olahraga balap MotoGP menjadi penting karena dapat meningkatkan daya tarik dari pariwisata Indonesia, meningkatkan ekspor, dan mendapatkan investor serta wisatawan manca negara, penyelenggaraan acara olahraga internasional di suatu negara telah menjadi strategi branding yang bertujuan untuk meningkatkan eksposur, reputasi, serta produk dari daerah tersebut. Tidak luput dari pengaruh positifnya saja, acara MotoGP ini sendiri memiliki sisi negatifnya tersendiri bagi beberapa kalangan masyarakat yang ada di Indonesia. Masyarakat banyak yang menyuarakannya melalui media social mengenai keberlangsungan dari pembangunan sampai pada saat acara diselenggarakan menimbulkan banyak reaksi positif dan negatif di media social, bahkan tidak jarang yang berpendapat untuk menolak pembangunan dan acara dimulai baik itu karena lokasi dan hal-hal lainnya.

Masyarakat banyak yang mengeluarkan pendapatnya pada media sosial Twitter mengenai penyelenggaraan MotoGP. Pandangan masyarakat dapat menjadi informasi bahwa masyarakat setuju atau tidak setuju dengan pembangun sirkuit Mandalika dan penyelenggaraan MotoGP di Lombok. Jika pandangan yang diberikan masyarakat positif atau baik maka dapat diasumsikan bahwa masyarakat setuju dengan penyelenggaraan kegiatan tersebut. Sebaliknya jika pandangan yang diberikan masyarakat negatif atau tidak baik maka dapat diasumsikan bahwa masyarakat tidak setuju dengan penyelenggaraan kegiatan tersebut. Dari berbagai macam pendapat tersebut dapat dilakukan analisis sentimen untuk melihat keberpihakan masyarakat terhadap penyelenggaraan MotoGP di Lombok.

## **PENGUMPULAN DATA**

Dilakukan pengumpulan data tweets dari situs Twitter menggunakan library snsrape.

Penelitian ini akan diimplementasikan menggunakan metode Deep Learning dengan arsitektur model Bidirectional Encoder Representations from Transformers (BERT). BERT menggunakan transformer, sebuah mekanisme attention yang mempelajari konteks antara kata-kata pada sebuah kalimat. Berlawanan dengan model satu arah yang hanya membaca teks secara sekuensial, karakteristik bidirectional pada BERT memungkinkan model untuk mempelajari konteks dari sebuah kata berdasarkan lingkungan. Model yang akan digunakan pada penelitian adalah model pre-trained

IndoBertTweet (Koto et al. 2021) yang merupakan model BERT yang dilatih dengan menggunakan dataset berbahasa Indonesia yang diambil dari situs Twitter untuk menganalisis sentimen penyelenggaraan MotoGP di Mandalika, Lombok.

## **METODE PENELITIAN**

Jika heading anda melebihi satu, gunakan level kedua heading seperti di bawah ini. Pengutipan dilakukan dengan mencantumkan nama penulis dan tahun penerbitan mengikuti kalimat. Penomoran harus diberikan pada gambar (gambar, grafik, foto dan peta), tabel dan persamaan matematika, rekasi kimia dan fisika berdasarkan urutannya.

### **Studi Literatur**

Dilakukan studi literatur dengan mengumpulkan berbagai referensi dan sumber seperti jurnal, artikel, buku, paper, situs-situs internet, dan seterusnya mengenai analisis sentimen, deep learning, BERT, dan lainlain

### **Pengumpulan Data**

Dilakukan pengumpulan data tweets dari situs Twitter menggunakan library snsrape.

### **Pelabelan Data**

Data akan diberikan label menjadi positif, netral, dan negatif.

### **Preprocessing Data**

Data akan dilakukan preprocessing sebelum dilatih untuk meningkatkan tingkat keakuratan model.

### **Pelatihan Model**

Pelatihan model mencakup aktivitas untuk melatih model IndoBERTTweet dan menyesuaikan parameter dan hyper parameter pada data yang sudah diolah sebelumnya.

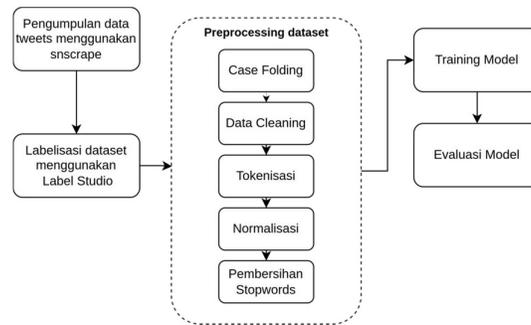
### **Evaluasi Model**

Mengevaluasi model yang sudah dilatih sudah memenuhi kriteria yang diharapkan.

## **HASIL DAN PEMBAHASAN**

### **Studi Literatur**

Metode yang digunakan untuk melakukan analisis sentimen terhadap penyelenggaraan MotoGP di Lombok dengan menggunakan teknik Bidirectional Encoder Representation from Transformers terdiri dari beberapa langkah berikut pada Gambar 1



Gambar 1. Arsitektur Umum

### Pengumpulan Data

Data yang diambil melalui situs twitter.com berupa cuitan-cuitan warganet mengenai pengadaan MotoGP di Lombok. Data dikumpulkan dengan menggunakan bantuan library Python yaitu, sncscrape. Proses pengumpulan data pada sncscrape menggunakan perintah sncscrape yang bisa dijalankan pada terminal atau cmd seperti pada Gambar 2

```
sncscrape \  
--jsonl \  
--max-results 10000 \  
twitter-search "MotoGP Mandalika" \  
> data-MotoGP.json
```

Gambar 2. Perintah sncscrape

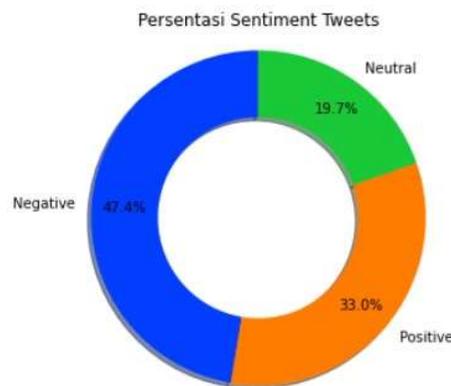
### Pelabelan Data

Analisis sentimen merupakan salah satu penerapan machine learning dengan berbasis supervised learning. Supervised learning membutuhkan data yang sudah memiliki label atau data yang sudah dianotasi. Pelabelan data ini diperlukan karena metode supervised learning membutuhkan contoh atau solusi dari setiap data yang akan dimasukkan ke dalam algoritma (Nasteski, 2017). Model algoritma akan dapat melihat, memahami, dan mengerti bagaimana komentar atau cuitan yang memiliki sentimen positif, netral, dan negatif. Contoh hasil pelabelan dapat dilihat pada Tabel 1.

**Tabel 1.** Contoh Hasil Pelabelan Datase

Teks	Label
Kalau gw sih lihat elu yang tolol. Buzzer pemecah belah bangsa. Gw sih dukung MotoGP mandalika dan Formula E Jakarta yg bikin Indonesia hebat bisa setara dgn negara lain. Merdeka	Positive
Makasi pak presiden jokowi. Sirkuit mandalika bikin kami bangga sebagai masyarakat, walaupun sampean tdk menang didaerah kami tp bagi sy engkaulah presiden yg betul2 bijaksana.	Positive
Yang pantas dikulik itu, Sirkuit Mandalika yang berbuah pisang. Lahan warga belum dibayar oleh pihak IDC.	Negatif
"Konon" Katanya Sirkuit Mandalika ngabisin Dana Hampir 2 Trilyun tp Lahan Warga kok nda Semuanya terbayar? Pada Kemana tu Duit larinya? Yg Lebih Aneh lagi kok "\$epertinya" KPK sama sekali tak Tertarik nyelidikin,Malahan lbh Sibuk Nguyek2 Dana Formula E yg cuma habis 150 Miliar?□	Negatif
kalau ada sepeda, mungkin bisa tuh start di sirkuit mandalika. siapa tahu dorna kedepannya ngadain balap sepeda roda tiga yang di dapat dari ayah, jangan lupa buat pencet bel "kring,,kring" ♦	Netral
Kabar Gembira! Tiket WSBK Mandalika Diskon 75 Persen + Bonus Jajal Sirkuit #Mandalika #WSBK	Netral

Berdasarkan proses pelabelan data didapatkan hasil data dengan sentimen negatif berjumlah 1427 data, sentimen netral berjumlah 607 data, dan sentimen positif berjumlah 1002 data yang dapat dilihat pada Gambar 3



**Gambar 3.** Perintah snsrape

**Preprocessing Data**

Preprocessing dilakukan untuk mengubah dataset menjadi dataset yang terstruktur untuk memastikan hasil analisis sentimen menjadi lebih baik. Tahapan preprocessing data terdapat beberapa tahapan yaitu case folding, data cleaning, tokenisasi, penghapusan stopword, dan normalisasi.

### Case Folding

Case Folding adalah tahap untuk membuat semua huruf pada teks atau kalimat menjadi huruf kecil. Tahap ini dilakukan untuk membuat semua karakter pada dataset menjadi sama, yaitu menggunakan huruf kecil sehingga kata “Motor” dan “motor” akan dianggap sama. Hasil dari proses *case folding* dapat dilihat pada Gambar 4.

	content	sentiment
0	pro ngaudit kerugian formula e. sklian audit m...	Neutral
1	@ch_chotimah2 @kpk_ri @kejaksaanri @st_burhanu...	Negative
2	@jackboea2 mandalika jadi ajang motogp terbaik...	Negative
3	indef merupakan lembaga independen yang ditunj...	Neutral
4	sirkuit mandalika siap gelar wsbk hingga motog...	Positive
...	...	...
620	trafik telkomsel melonjak 96 persen selama mot...	Neutral
621	para rider motogp minta mandalika diperbaiki l...	Positive
622	total penonton di motogp mandalika capai 102.8...	Positive
623	indonesia mampu bersaing <a href="https://t.co/pdy10zr0tu">https://t.co/pdy10zr0tu</a>	Positive
624	#tolakboostersyaratmudik	Neutral

Gambar 4. Data setelah dilakukan *Case Folding*

### Data Cleaning

Tokenisasi adalah suatu proses yang dilakukan untuk memecah kalimat menjadi potongan kata, tanda baca, dan ekspresi bermakna lainnya sesuai dengan ketentuan bahasa yang digunakan. Salah satu cara untuk meng tokenisasi teks dalam bahasa *Python* adalah dengan menggunakan *library* NLTK (*Natural Language Toolkit*). *Library* NLTK menyediakan *function word\_tokenize* untuk memecah kalimat menjadi kata-kata. Berikut ini adalah perbandingan hasil data yang asli dengan data yang sudah dibersihkan dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Hasil *Data Cleaning*

Sebelum	Sesudah
@nusantara_one @serunipuspaalam @yennywahid aturan jangan ke wan toa! gelaran macam formula e yang tarik ulur pementasannya. mending tanya gelaran internasional macam motogp mandalika. jelas2 diakui dunia atas kesuksesannya. bertanyalah ke expert	aturan jangan ke wan toa gelaran macam formula e yang tarik ulur pementasannya mending tanya gelaran internasional macam motogp mandalika jelas2 diakui dunia atas kesuksesannya bertanyalah ke expert jangan ke mr bullshit grinning_face grinning_face

### Tokenisasi

Case Folding adalah tahap untuk membuat semua huruf pada teks atau kalimat menjadi huruf kecil. Tahap ini dilakukan untuk membuat semua karakter pada dataset menjadi sama, yaitu menggunakan huruf kecil sehingga kata “Motor” dan “motor” akan dianggap sama. Perbandingan data yang asli dengan data yang sudah ditokenisasi dapat dilihat pada Tabel 4.

**Tabel 4.** Perbandingan Hasil Tokenisasi

Sebelum	Sesudah
@nusantara_one @serunipuspaalam @yennywahid aturan jangan ke wan toa! gelaran macam formula e yang tarik ulur pementasannya. mending tanya gelaran internasional macam motogp mandalika. jelas2 diakui dunia atas kesuksesannya. bertanyalah ke expert	aturan jangan ke wan toa gelaran macam formula e yang tarik ulur pementasannya mending tanya gelaran internasional macam motogp mandalika jelas2 diakui dunia atas kesuksesannya bertanyalah ke expert jangan ke mr bullshit grinning_face grinning_face

### Normalisasi

Normalisasi dilakukan untuk mengubah kata-kata tidak baku menjadi kata yang baku atau sesuai dengan ejaan. Hal ini dilakukan karena pada umumnya banyak sekali yang menggunakan kata-kata gaul seperti: yg, okeh, bnr, pngn, km, mantep, manteppppp, dan lain-lain. Jika kata-kata tersebut tidak melewati proses normalisasi, maka sistem akan menganggap kata mantep, mantepppp, mantepppppp, mantap adalah kata yang berbeda yang seharusnya adalah kata yang bermakna sama. Perbandingan data yang asli dengan data yang sudah dinormalisasi dapat dilihat pada Tabel 5.

**Tabel 5.** Perbandingan Hasil Normalisasi

Sebelum	Sesudah
cuma 5m doang keuntungannya....dana yg keluar 500m.... kalo kata kadrun keuntungan formula e melibihi motogp mandalika... kadrunn somplak...!!!!	cuma 5m doang keuntungannya dana yg keluar 500m kalo kata kadrun keuntungan formula e melibihi motogp mandalika kadrun somplak

### Penghapusan Stopwords

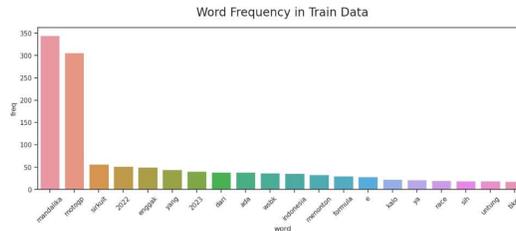
*Stopwords* adalah kata-kata yang kurang memiliki makna. Penghapusan kata-kata *stopwords* akan mengubah kata yang memiliki informasi yang kurang penting sehingga menyisakan kata yang memiliki informasi yang lebih penting. Tahap ini akan menggunakan *library stopwords* Bahasa Indonesia yang disediakan oleh NLTK. Perbandingan data yang asli dengan data yang sudah dihapus *stopwords* dapat dilihat pada Tabel 6.

**Tabel 6.** Perbandingan Hasil Penghapusan *Stopwords*

Sebelum	Sesudah
tapi ngapain anjir ke mandalika lombok jangan bikin berharep lu pada gada apa apa lu mau nonton motogp? dh enggak ada. apa lu mau lomba motogp ngab ato mau ngeliat patung puteri mandalika lu pada enggak mungkin anjir ke mandalika soalnya masih banyak pulau yang lebih bagus mana panas lagi	tapi ngapain anjir mandalika lombok bikin berharep lu pada gada apa apa lu mau nonton motogp dh. apa lu mau lomba motogp ngab ato mau ngeliat patung puteri mandalika lu pada mungkin anjir mandalika soalnya masih banyak pulau lebih bagus mana panas lagi

### HASIL PREPROCESSING DATA

Hasil dari preprocessing menghasilkan korpus yang berjumlah sebesar 1682 kata. Frekuensi kata yang terbanyak ditempati oleh kata mandalika dan motogp dan diikuti oleh kata yang berkaitan dengan terselenggaranya MotoGP Mandalika yang dapat dilihat pada Gambar 5.



Gambar 5. Frekuensi Kata dalam Data Training

Hasil *wordcloud* menunjukkan kata terbanyak pada korpus yang juga ditempati oleh kata motogp dan mandalika dan diikuti oleh kata yang berkaitan dengan terselenggaranya MotoGP Mandalika Lombok yang dapat dilihat pada Gambar 6.



Gambar 6. Wordcloud Training Data

### PELATIHAN MODEL

Model BERT yang akan dipakai pada penelitian ini adalah model IndoBERTTweet dari IndoLEM. Model ini dipilih karena model tersebut telah dilatih menggunakan empat ratus juta kata dalam bahasa Indonesia yang diambil dari situs Twitter (Koto et al., 2021). Penelitian ini menggunakan library Transformers yang disediakan oleh HuggingFace yang akan diimplementasikan menggunakan framework TensorFlow dari Google. Library Transformers menyediakan ribuan pre-trained model dalam ratusan bahasa yang dapat digunakan untuk 12 downstream tasks seperti klasifikasi, ekstraksi informasi, translasi, tanya jawab, summarization, text generation dan lain-lain. Transformers juga didukung oleh banyak framework machine learning dan deep learning seperti PyTorch, JAX, dan TensorFlow.

Sebelum dilakukannya pelatihan model, dataset harus disesuaikan dengan input yang dapat diterima oleh model BERT. Tahapan yang akan dilakukan sebelum melatih model diantaranya memuat dan memproses dataset sehingga dataset dapat ditokenisasi ke dalam bentuk input example. Selanjutnya dilakukan tokenisasi menggunakan fungsi khusus dari library Transformers sehingga dataset dapat dimasukkan ke dalam model BERT. Tahap berikutnya adalah memuat model dari library

Transformers dan melatih model dengan dataset yang sebelumnya sudah dipersiapkan. Tahap terakhir adalah mengevaluasi hasil pelatihan model dari beberapa metrik seperti akurasi, validasi akurasi, loss, dan validasi loss.

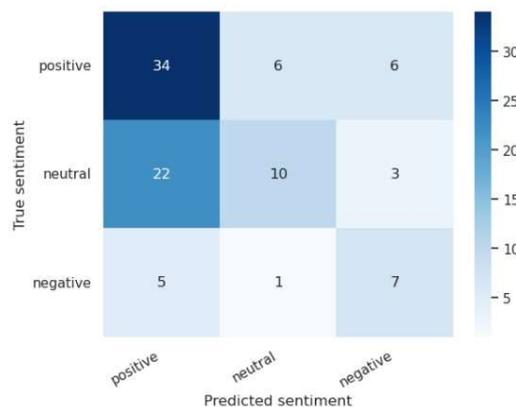
### EVALUASI MODEL

Evaluasi model akan dilakukan terhadap test dataset dengan menggunakan confusion matrix untuk mengukur performa model yang didapatkan dari proses fine-tuning. Hasil yang akan dihitung adalah precision, recall, accuracy dan f1- score. Kode pada Gambar 3.50 akan melakukan tokenisasi pada test dataset yang selanjutnya akan di prediksi dengan model IndoBERTTweet. Hasil output dari model merupakan logits yang harus dihitung dengan softmax function untuk menghasilkan hasil prediksi kumulatif dari test dataset. Hasil dari softmax function selanjutnya dilakukan function argmax untuk melihat nilai terbesar dari suatu array atau logits yang sudah diprediksi sebelumnya. Apabila nilai dari argmax berupa 0 maka kalimat tersebut bersentimen positif, apabila bernilai 1 maka kalimat tersebut bersentimen netral, dan apabila bernilai 2 maka kalimat tersebut bersentimen negatif.

Dari confusion matrix pada Gambar 7 dapat disimpulkan bahwa terdapat:

- a. True Positive sebesar 34 data.
- b. True Neutral sebesar 10 data.
- c. True Negative sebesar 7 data.
- d. False Positive sebesar 31 data.
- e. False Neutral sebesar 27 data.
- f. False Negative sebesar 39 data.

Dapat terlihat bahwa model cukup seimbang dalam memprediksi keseluruhan data. False Negative memiliki kesalahan yang lebih besar dibandingkan sentimen yang lain karena data yang bersentimen negatif cenderung lebih banyak dibandingkan data bersentimen netral dan positif.



**Gambar 7.** Confusion Matrix dari Evaluasi Model

Setelah mendapatkan hasil dari confusion matrix, perhitungan accuracy, precision, recall, dan f1-score dilakukan menggunakan fungsi classification\_report dari library Scikit-learn.

Perhitungan evaluasi menggunakan fungsi `classification_report` dari Scikit-learn dapat dilihat pada Gambar 8.

	precision	recall	f1-score	support
positive	0.56	0.74	0.64	46
neutral	0.59	0.29	0.38	35
negative	0.44	0.54	0.48	13
accuracy			0.54	94
macro avg	0.53	0.52	0.50	94
weighted avg	0.55	0.54	0.52	94

**Gambar 8.** *Classification Report* Hasil Evaluasi Model

Dari perhitungan manual dan perhitungan menggunakan fungsi `classification_report` didapatkan hasil yang sesuai menurut confusion matrix. Pada Gambar 3.53 dapat disimpulkan bahwa model mendapat skor cukup tinggi pada sentimen positif, dan skor lebih rendah pada sentimen netral dan negatif.

## KESIMPULAN DAN SARAN

### Kesimpulan

Analisis sentimen mengenai pelaksanaan MotoGP di Lombok berhasil dilakukan menggunakan model pre-trained IndoBERTTweets dengan data yang diambil dari situs Twitter dari tanggal 01-01-2022 sampai 30-10-2022. Pengambilan data menghasilkan 3036 data tweets mengenai MotoGP Mandalika dengan bersentimen negatif berjumlah 1427 data, data bersentimen netral berjumlah 607 data, dan data bersentimen positif berjumlah 1002 data.

Pengklasifikasian dengan metode BERT dilakukan dengan pemilihan hyperparameters yaitu batch size sebesar 32, optimizer menggunakan Adam dengan learning rate  $3e-6$  atau 0.000003, dan epoch 25. Hasil evaluasi dari model tersebut mendapatkan akurasi sebesar 55%. Precision untuk positif sebesar 56%, netral sebesar 59%, dan negatif sebesar 44%. Recall untuk positif sebesar 74%, netral sebesar 29%, dan negatif 54%. F1-score untuk positif sebesar 64%, netral sebesar 38%, dan negatif sebesar 48%.

Berdasarkan hasil-hasil pengujian menunjukkan bahwa model dapat mengklasifikasikan sentimen positif dengan baik, dan kesulitan untuk mengklasifikasi sentimen netral dan negatif. Hasil tersebut juga menunjukkan bahwa masyarakat memiliki opini positif terhadap penyelenggaraan MotoGP di Lombok.

### Pelatihan Model

Saran yang dapat dilakukan untuk penelitian selanjutnya adalah dengan menggunakan dataset yang lebih banyak agar mendapatkan akurasi yang lebih baik, pelabelan dilakukan oleh ahli bahasa agar kesalahan dalam pelabelan dapat diminimalisir, melakukan penyebaran perangkat lunak agar model dapat bisa dicoba secara langsung.

## DAFTAR PUSTAKA

- Alammar, J. (n.d.). The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning). URL: <https://jalammar.github.io/illustrated-bert/> [Tanggal Akses 20 Juli 2022]
- Aliyah Salsabila, N., Ardhito Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2018, November). Colloquial Indonesian lexicon. 2018 International Conference on Asian Language Processing (IALP). URL: <http://dx.doi.org/10.1109/ialp.2018.8629151>
- Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 21(4), 3039–3071. URL: <https://doi.org/10.1109/comst.2019.2926625>
- Chollet, F. (2021). *Deep Learning with Python*, second edition. Simon and Schuster.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, October 11). BERT: Pre-training of deep bidirectional transformers for language understanding. ArXiv.Org. URL: <https://arxiv.org/abs/1810.04805>
- Eliasy, A. (2020, July 1). Fig. 2. The comparison between Recurrent Neural Network (RNN) and... Research Gate. URL: [https://www.researchgate.net/figure/The-comparison-between-Recurrent-Neural-Network-RNN-and-Feed-Forward-Neural-Network\\_fig1\\_338672883](https://www.researchgate.net/figure/The-comparison-between-Recurrent-Neural-Network-RNN-and-Feed-Forward-Neural-Network_fig1_338672883)
- Gärtler, M., Khaydarov, V., Klöpffer, B., & Urbas, L. (2021). The machine learning life cycle in chemical operations – status and open challenges. *Chemie Ingenieur Technik*, 93(12), 2063–2080. URL: <https://doi.org/10.1002/cite.202100134>
- Géron, A. (2019). *Hands-On machine learning with Scikit-Learn, Keras, and TensorFlow*, 2nd edition.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*. URL: <https://doi.org/10.1007/s11042-022-13428-4>
- Koto, F., Lau, J. H., & Baldwin, T. (2021, September 10). IndoBERTweet: A pretrained language model for indonesian twitter with effective domain- specific vocabulary initialization. ArXiv.Org. URL: <https://arxiv.org/abs/2109.04607>
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS.B*, 4, 51–62. URL: <https://doi.org/10.20544/horizons.b.04.1.17.p05>
- Saleh, Z. (2019, April 9). Artificial intelligence definition, ethics and standards. URL: [https://www.researchgate.net/publication/332548325\\_Artificial\\_Intelligence\\_Definition\\_Ethics\\_and\\_Standards](https://www.researchgate.net/publication/332548325_Artificial_Intelligence_Definition_Ethics_and_Standards)
- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, May 14). How to fine-tune BERT for text classification? ArXiv.Org. URL: <https://arxiv.org/abs/1905.05583>
- Yoo, B., & Rayz, J. T. (2021). Understanding emojis for sentiment analysis. *The International FLAIRS Conference Proceedings*, 34. URL: <https://doi.org/10.32473/flairs.v34i1.128562>